Passive Devices for Terahertz Frequencies

Timothy D. Drysdale, BE (1^{st} *Class Hons.*)

A thesis presented for the degree of
Doctor of Philosophy
in
Electrical and Electronic Engineering
at the
University of Canterbury,
Christchurch, New Zealand.

22 January 2003

ABSTRACT

Terahertz technology is a relatively new field of electromagnetic study and interest is rapidly growing in the wake of dramatic imaging demonstrations. Other applications are expected to follow, and they will need passive devices with functionality already found in more familiar microwave and visible regions of the electromagnetic spectrum, but presently missing in the terahertz region. Two fundamental devices in particular are variable polarisation compensators, and tunable frequency-selective filters.

This work represents the first demonstration of a variable polarisation compensator using subwavelength patterned features (artificial dielectrics). Following on from the original proposal, this work contains a complete and thorough investigation including the development of a bulk silicon micromachining fabrication process, full characterisation of the device performance in the W-band (70 – 110GHz) and comprehensive simulations of the device, including detailed simulation of three distinct new designs with improved performance (continuously-variable retardance with maximum in excess of quarter- and half wave). The third of the three designs is capable of extremely low insertion loss (< 0.6 dB) and overcomes a difficulty of the original design that prevented zero retardance in a practical device.

Secondly, a new tunable photonic crystal filter is proposed and demonstrated. Easily accessible external control surfaces integrated into the interlocking plates of a layer-by-layer photonic crystal allow unprecedented contol over the number and type of defects within the structure, all of which may be tuned "on-the-fly". Devices are initially investigated with a full-vector electromagnetic finite-difference time-domain technique, to reveal the influence of the design dimensions on the band gap as well as the effect of the defects. A two-plate metal device having four layers of rods is constructed and measured in the W-band. In good agreement with the simulations, it is experimentally determined that a moveable passband is centered at 81 GHz, with a quality factor of 11, and a tuning shift of 1.7 GHz for a plate movement of 450 μm.

PREFACE

This dissertation describes research undertaken in the Department of Electrical and Electronic Engineering at the University of Canterbury (New Zealand) and in the Department of Electronics and Electrical Engineering at the University of Glasgow (United Kingdom) between March 1999 and December 2002. I am grateful to my supervisor R. J. Blaikie and our collaborator D. R. S. Cumming for suggesting the topic of research and maintaining a keen interest throughout.

The following papers and patents were prepared during the course of this work:

T.D. Drysdale, R.J. Blaikie, D.R.S. Cumming, "Calculated and measured transmittance of a tuneable metallic photonic crystal filter for terahertz frequencies," *Applied Physics Letters*, accepted for publication in the December 29, 2003 issue[1].

T.D. Drysdale, R.J. Blaikie, H.M.H. Chong, D.R.S. Cumming, "Artificial dielectric devices for variable polarization compensator at millimeter and submillimeter waves", *IEEE Transactions on Antennas and Propagation*, vol. 51, no. 11, pp. 3072 – 3079, 2003.

T.D. Drysdale, G. Mills, S.M. Ferguson, R.J. Blaikie, D.R.S. Cumming, "Terahertz tuneable filters made by self-releasing deep dry etch process", *Microelectronic Engineering*, Cambridge, U.K., September 2003, accepted for publication[1].

T.D. Drysdale, G. Mills, S.M. Ferguson, R.J. Blaikie, D.R.S. Cumming,"Metallic tunable photonic crystal filter for TeraHertz frequencies", *J. Vac. Sci. Technol. B*, Tampa, Florida, May 27 – 30, 2003, accepted for publication.

T.D. Drysdale, R.J. Blaikie, D.R.S. Cumming, "Tuneable photonic crystal filter for terahertz frequency applications", *Proc. SPIE*, vol. 5070, pp. 89 - 97, Orlando, Florida, April 21 – 25, 2003.

T.D Drysdale, R.J. Blaikie, D.R.S. Cumming, "Filter Device", United Kingdom Patent Application and Patent Cooperation Treaty International Application, first priority GB 0126737.6, 7 Nov 2001.

[1]This paper prepared after submission, whilst employed as a Research Assistant to D. R. S. Cumming at the University of Glasgow

T.D. Drysdale, H.M.H. Chong, R.J. Blaikie, D.R.S. Cumming, "A variable polarisation compensator using artificial dielectrics for millimetre and submillimetre waves", *Electronics Letters*, vol. 37, no. 3, pp. 149 – 150, 2001.

K.K. Ma, T.D. Drysdale, R.J. Blaikie, D.R.S. Cumming., "Novel silicon bulk micromachining process for submillimeter rectangular waveguide fabrication", *Proc. SPIE*, vol. 4407, pp. 372 – 379, Edinburgh, Scotland, May 30 – June 1, 2001.

T.D. Drysdale, H.M.H. Chong, R.J. Blaikie, D.R.S. Cumming, "Micromachined submillimeter & millimeter wave variable polarisation compensator", *Technical digest of International Microwave Symposium*, pp. 535 – 538, May 20 – 25, Phoenix, Arizona, USA, 2001.

R.J. Blaikie, T.D. Drysdale, H.M.H. Chong, D.R.S. Cumming, "Artificial dielectric materials for mm and sub-mm wavelengths", *Workshop on Smart Materials: Technological Applications*, April 9-10, Dunedin, New Zealand, 2001.

T.D. Drysdale, R.J. Blaikie, D.R.S. Cumming, "A Variable Artificial Dielectric Retarder for TeraHertz Frequencies", *Proc. 6th ENZCon*, Auckland, New Zealand, pp. 146 – 150, 1999.

ACKNOWLEDGEMENTS

First and foremost I thank R.J. Blaikie for supervising this study. His enthusiasm, guidance, and support throughout have been much appreciated. Thanks are also due to D.R.S. Cumming, who has been the source of many useful ideas and organised access to facilities at the University of Glasgow. Together, R.J.B. and D.R.S.C. have made this study a rewarding experience.

I would like to acknowledge financial support from a University of Canterbury Doctoral Scholarship and research funding from the Marsden Fund of the Royal Society of New Zealand (Contract UOC-803). Towards the end of the Ph.D. I moved to the U.K. to take up a position as a research assistant to D. R. S. Cumming, funded by the Scottish Enterprise's Proof of Concept scheme.

Technical staff at both universities were instrumental in fabricating experimental equipment, teaching fabrication processes and keeping labs running, especially H. Devereaux, G. Turner, D. Sallis, S. Downing, P. Lambert (Canterbury), K. Piechowiak, J. Kelly, S. Ferguson, A. Ross (Glasgow). Fellow NEST group students assisted greatly, in particular S. J. McNab, E. D. Walsby, and K. K. Ma who showed me the ropes, B. R. Rong and G. Kanesan who shared the office, and S. Drake, D. Lee, V. Christie and M. Konijn who read chapters. NEST group staff M. Alkaisi and R. Cheung also provided a helping hand from time to time. Members of the Electrical wing read chapters (T. Cussins), and provided necessary diversions and entertaining discussions (L. Toorenburg, P.J. Barclay, H.J. Callow, C. Forne, S. Hardie, E. Pilbrow). R. Munro forwarded my mail, H. Fielding looked after paperwork, and along with P. Haigh and B. Hall, kept me informed about life on the fourth floor.

In Glasgow, H.M.H. Chong "consulted" on I. Thayne's W-band measurement system. H.M.H.C. was assisted in giving me a warm welcome to Glasgow by C. Walker, C. Young, S. Hamill, S.S. Kim, C. Hadyoon, E. Boyd, F. McEwan, D. Morgan, P. Hammond, and G. Mills along with the basketballers, members of the MST group and the basement coffee club. C. Walker, G. Burns and G. Irvine read chapters. M. Telfer and K. Telfer were understanding flatmates.

Thanks are also due to A. Bertel for his karate training, to B. Martin for his advice on all things, and last but not least to my family, Peri, Alex and Emily, for their love and support throughout.

NOTATION

(xyz)	individual plane, or surface plane of wafer depending on context
$\{xyz\}$	group of (xyz)-equivalent planes
$[xyz]$	individual crystallographic direction - $[xyz]$ is normal to the plane (xyz)
$\langle xyz \rangle$	group of $[xyz]$-equivalent directions
c	speed of light in a vacuum, 2.998×10^8 m/s
$\mathbf{E}$	electric field vector
F	Farad
hr	hour
$\mathbf{H}$	magnetic field vector
H	Henry, or magnitude of magnetic field depending on context.
k	wavenumber
$\mathbf{k}$	wavevector
$\mathbf{K}$	grating vector
m	metre
min	minute
n	refractive index
$\tilde{n}$	complex refractive index ($\tilde{n} = n + \kappa$)
s	second
ε_0	permittivity of free space, 8.854×10^{-12} F/m
ε_r	relative permittivity, $\varepsilon_r = \varepsilon/\varepsilon_0$
η_0	intrinsic impedance of free space, 376.6 Ω
κ	extinction coefficient of optical material
λ	wavelength
μ_0	permeability of free space, $4\pi \times 10^{-7}$ H/m
ξ	ellipticity
Γ	retardance
Λ	period
Ω	resistance or impedance, Ohms

GLOSSARY

AC	Alternating Current
AR	Anti-Reflection (coating)
DARTH-VADR	Dual Axis Rectangular Tooth Half-wave VADR
DI	De-Ionised water
DRIE	Deep Reactive Ion Etch
DUT	Device Under Test
EMT	Effective Medium Theory
FCT	Face Centred Tetragonal (lattice)
FDTD	finite difference time domain
FWHM	Full Width Half Maximum
JPEG	Joint Photographic Experts Group
LRL	Line Reflect Line (VNA calibration)
PBG	Photonic Band Gap
RCWA	Rigorous Coupled-Wave Analysis
RF	Radio Frequency
RIE	Reactive Ion Etch
R-VADR	Rectangular-groove VADR
R^2-VADR	double Rectangular-groove VADR
SC	Simple Cubic (lattice)
SEM	Scanning Electron Microscope
ST	Simple Tetragonal (lattice)
TE	Transverse Electric
TEMPEST	FDTD code from University of California, Berkeley
THz	TeraHertz (10^{12} Hz)
TM	Transverse Magnetic
VADR	Variable Artificial Dielectric Retarder
VNA	Vector Network Analyser
VPBGF	Variable Photonic Band Gap Filter
W-band	Band of frequencies between 70 – 110 GHz

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 INTRODUCTION

Terahertz (THz) frequency research is one of the most exciting fields of electromagnetic research currently under investigation. Recent demonstrations of optoelectronic "T-ray" imaging systems have been responsible for much of this excitement, due to the dramatic pictures they have produced, including a map of the water distribution in a leaf, the temperature distribution in a flame and the layout of packaged semiconductors [1, 2]. While it is true that terahertz technology has been held back until recently by a lack of suitable sources and detectors, laboratory systems employing ultrafast optical pulses to stimulate emission of terahertz radiation are now well understood.

Furthermore, a great effort is being expended worldwide in developing novel sources and detectors so that the needs of future terahertz systems may be met. Such systems, with spectroscopic, imaging, or communications functions, may need to be portable in order to offer a vehicle for the widespread adoption of terahertz technology. Due to unique combinations of properties of radiation at terahertz frequencies, these three basic functionalities encompass a diverse and numerous range of applications (biomedical imaging [3], early-warning detection of chemical and biological warfare agents [4], determination of the fat content in packaged food [5], secure communications between satellites, and the search for forming stars and life-supporting planets [6] to name but a few). However, if a system is to perform any function other than the straightforward imaging demonstrated to date (if anything at these frequencies may be called straightforward), then additional devices will be required, just as they are in other frequency regimes. Unfortunately, like conventional sources and detectors, the passive devices that have been developed for the other frequencies, do not always transfer well to terahertz frequencies. Therefore, it is the purpose of this study to investigate a number of important passive devices for use at terahertz frequencies.

The particular devices investigated here are variable polarisation compensators using artificial dielectrics, and variable spectral filters using photonic bandgap structures. Details of the devices are presented in chapter 2. Schematic pictures of the two

classes of device studied here are presented without detailed explanation in Figure 1.1 for the interest of the reader who may already be familiar with the general subject area of structured materials. The double-ended arrows indicate the direction in which the plates are moved to obtain variable characteristics, and the large outline arrows indicate the direction of the wave propagation.

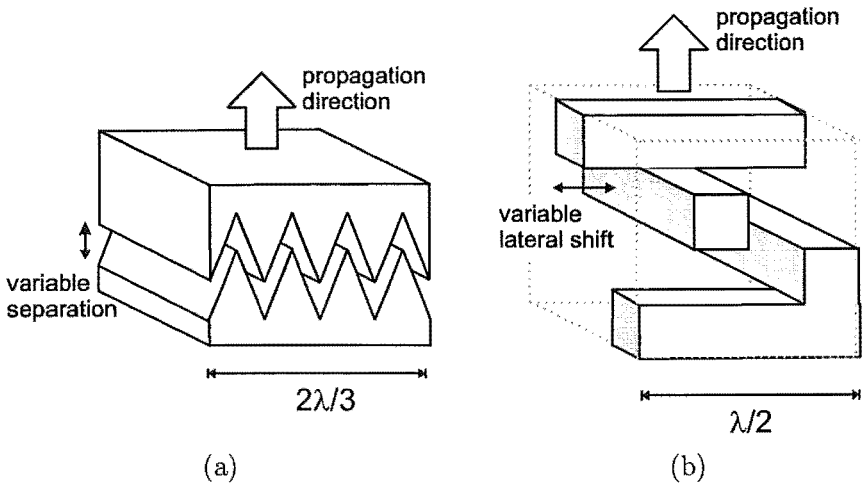


Figure 1.1 Two devices representative of those studied in this thesis: (a) variable polarisation compensator constructed from two V-grooved silicon plates that vary in separation (four periods shown); (b) variable photonic band gap filter constructed from two plates, each having an orthogonal grid (one period shown), that shift laterally with respect to each other. The small double-ended arrows show the direction of plate movement that achieves variability, while the large outline arrows indicate the wave propagation directions. The approximate size relative to the design wavelength λ is also indicated.

The polarisation compensator shown in Fig 1.1(a) is made of two silicon plates patterned with interlocking V-grooves, with four periods shown. Since silicon has little loss at terahertz frequencies, electromagnetic waves may propagate through the device. The V-groove gratings of the polarisation compensator are periodic in only one direction, thus the optical path length through the device is different for incident radiation with its electric field vector oriented parallel the gratings, as compared to radiation with the electric field oriented perpendicular to the gratings (birefringence). The exact difference in path length is controlled by the separation of the plates, and consequently the relative change in the phase of the two electric field components may also be controlled, allowing the polarisation state to be easily manipulated.

The filter shown in Fig 1.1(b) is constructed from a stack of two or more dielectric or metallic plates, each having two orthogonal grids; only one unit cell of a device having two plates is shown in this case. The characteristic dimensions of the filter are larger than that of the polarisation compensator. The gratings of the photonic bandgap filter form a periodic structure, giving rise to interfering reflections of the electromagnetic waves within the device. For certain bands of frequencies, the interference is destructive, and the electromagnetic waves may not propagate (bandgap). By shifting the plates

either laterally or vertically with respect to each other, the periodicity of the crystal is disturbed, altering the band of wavelengths for which electromagnetic radiation may pass through the crystal. The metallic versions of the device are completely different to metal mesh etalon filters used in the infrared [7], partly because their only movement is lateral.

The next section of this chapter provides a brief overview of terahertz frequencies. The sections thereafter set out the motivation for choosing the particular devices, and the aims of this study. The final section outlines the organisation of the rest of this thesis.

1.2 TERAHERTZ FREQUENCIES

This section provides a brief overview of selected highlights from the field of terahertz technology. For a review that is far more thorough than possible within the present space constraints, please see reference [6].

1.2.1 Properties

The terahertz region corresponds to frequencies of about 100 GHz - 10 THz, or wavelengths of 3mm – 30μm [8]. The relationship between THz frequencies and the surrounding electromagnetic spectrum is illustrated in Figure 1.2. At the lower end, the THz region overlaps the millimetre-wave region, while at the upper end it overlaps the far infrared. It is this overlap between 'electronic' frequencies and 'optical' frequencies that is one of the reasons that makes terahertz research interesting and challenging.

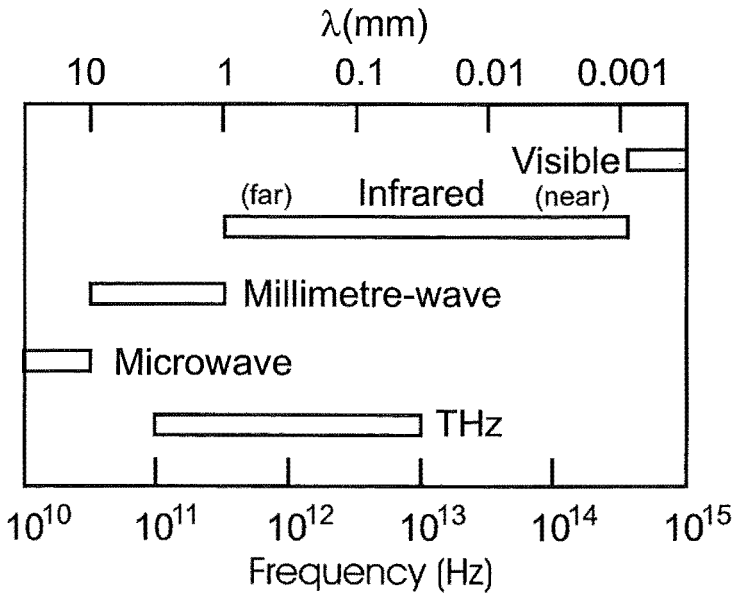


Figure 1.2 Electromagnetic spectrum in the vicinity of the THz region.

The photon energies range from 0.41 meV at 100 GHz, to 41 meV at 10 THz [9]. These energies are insufficient to cause electron transitions in matter, and thus radiation at THz frequencies may be considered non-ionising. This is important, because ionising radiation (energy $> 2 \sim 3eV$) can cause severe damage to biological tissue, due to the creation of highly-reactive transient species [10]. So long as thermal effects are avoided by using low average powers, biological tissue may be safely irradiated with terahertz radiation without the strict dosage limits that must be imposed for X-rays [11].

These moderate photon energies correspond to the energies of rotational, torsional and vibrational state transitions in a wide range of molecules [12]. This makes terahertz attractive for spectroscopic applications [13, 14]. It is a mixed blessing that terahertz frequencies are strongly absorbed in water. On one hand, long distance transmissions (>1km) are generally impossible given the water content of the atmosphere, and the associated absorption spectrum [6]. On the other hand, communications are more secure due to the containment of stray signals, and in densely populated areas, frequency reuse is more efficient. Additionally, non-invasive water content measurements may be made more accurately than ever before [1]. Other polar substances are absorbing, whereas non-polar substances are expected to be transmissive to some extent, for example crystals such as sapphire, silicon and germanium [15]. In fact, intrinsic silicon is an excellent optical material in the terahertz region, with low loss and low dispersion, as illustrated in Figure 1.3 [16].

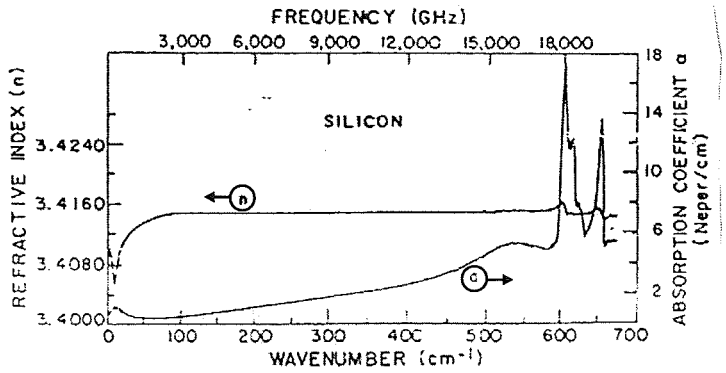


Figure 1.3 Refractive index and absorption coefficient for silicon, at terahertz frequencies. Reproduced from [16].

1.2.2 Sources and Detectors

Historically, it has been difficult to generate coherent THz radiation in a compact, efficient, reliable, and inexpensive manner, so activities are restricted to much larger installations. Where electron accelerators are available, and if Gigawatt powers and tunability are required, free electron lasers may be used [17]. On a smaller scale,

suitable for individual laboratories, ultrafast-optical-pulse-stimulated emission may be used.

Two main ultrafast-optical-pulse-stimulated emission systems have been developed, and both are used to create freely propagating, broadband pulses of terahertz radiation with energies in the range of several nano-watts to tens of micro-watts [5]. Although not the first method proposed, photoconductive switch-based systems were the first to be used widely, and were well suited to terahertz time domain spectroscopy [18–21]. In this method, ultrafast (sub-picosecond duration) optical pulses (typically $\lambda = 800\text{nm}$) are focused onto a silicon photoconductive switch. This generates carriers in the switch substrate, that are then accelerated by a constantly-applied DC-bias causing terahertz radiation to be emitted.

A typical setup, where the photoconductive switch is placed between the two arms of a dipole antenna, is illustrated schematically in Figure 1.4. In order to obtain coherent detection, the optical pump pulses were split and focused onto a similar photoconductive switch, located between the arms of a second dipole antenna. By varying the delay of the pulse at the receiving antenna, it was possible to obtain a time-resolved recording of the received terahertz pulse with an excellent signal to noise ratio.

A related technique involves sweeping the beam along a coplanar waveguide [22,23] to create field patterns that matched the propagating TEM mode. Sometimes an additional impedance-matched antenna [24, 25], sapphire lens [26], or spherical metal mirror [27] was added.

The second method, using optical rectification at the emitter, and electro-optic sampling via the Pockels effect at the detector, is the favoured system for real-time 2-dimensional imaging systems, because of the speed at which data may be collected [28–35]. In this method, various crystals with high second order susceptibilities ($\chi^{(2)}$) may be used for the emitter, notably including the electro-optic crystals zinc telluride (ZnTe), lithium niobate ($LiNbO_3$) (which may be used in the detector) [30,36,37], and gallium arsenide (GaAs) [38]. Terahertz frequency radiation is emitted from the crystal as a result of difference frequency generation, caused by the incident ultrafast optical pulse. The magnitude of the emitted radiation, E_{THz}, is proportional to the induced second order polarisation[1], P, which may be calculated in the time domain by

$$P^2_{THz}(t) = \varepsilon_0 \chi^{(2)} E(t) E^*(t) \tag{1.1}$$

where ε_0 is the permittivity of free space, and $E(t)$ and its conjugate $E^*(t)$ represent the electric field of the incident optical pulse. The emitted terahertz pulse represents the intensity envelope of the incident optical pulse, as illustrated schematically in Fig-

[1]refers to the dipole moment density of the atoms or molecules in the crystal; not to be confused with the unrelated concept of the polarisation of an electromagnetic wave, to which extensive reference is made elsewhere in this thesis.

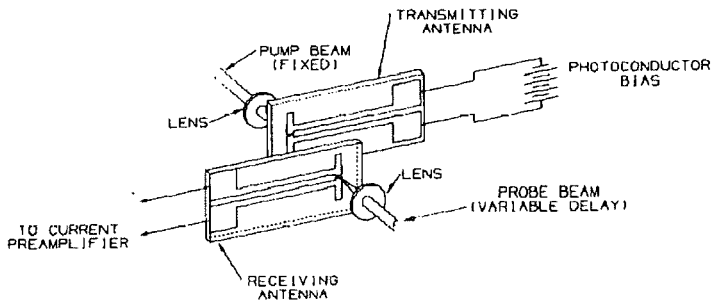


Figure 1.4 Photoconductive generation of THz broadband pulse using silicon photoconductor on sapphire. The antenna face each other, separated by 2mm, and are illuminated from behind, through the substrate. (Reproduced from [21] Fig. 2).

ure 1.5. Higher THz frequencies can be produced by shortening the incident optical pulse width, since the maximum emitted frequency corresponds to the bandwidth of the incident optical pulse, typically in the range of 2 THz for current systems (with optical pulse duration $\sim$ 250 fs). However, the technique has been demonstrated to work at frequencies up to, and beyond 30 THz [39].

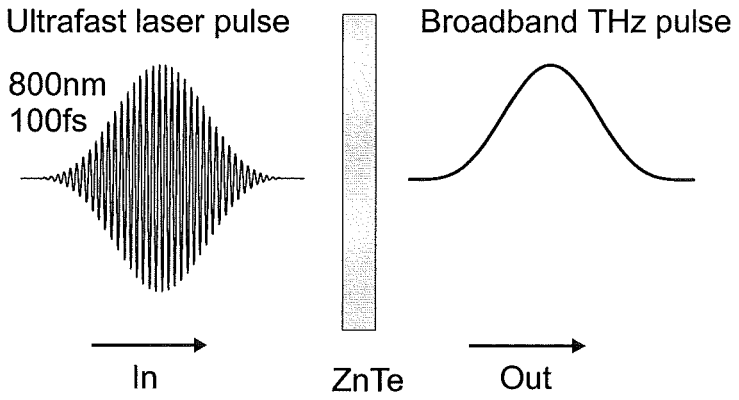


Figure 1.5 Optical rectification in a zinc telluride (ZnTe) crystal converts input ultrafast laser pulse into a broadband THz pulse.

The detection follows the same scheme as before, in that a co-propagating variably delayed optical pulse is used to probe the detector as shown in Figure 1.6, although in this case it is a electro-optic crystal. The received terahertz radiation alters the polarisation of the electro-optic sampling crystal, and that changes the polarisation state of the optical pulse. The polarisation change in the optical pulse may be determined electronically by splitting the signal with a Wollaston polariser and detecting the two components with a balanced photo-detector.

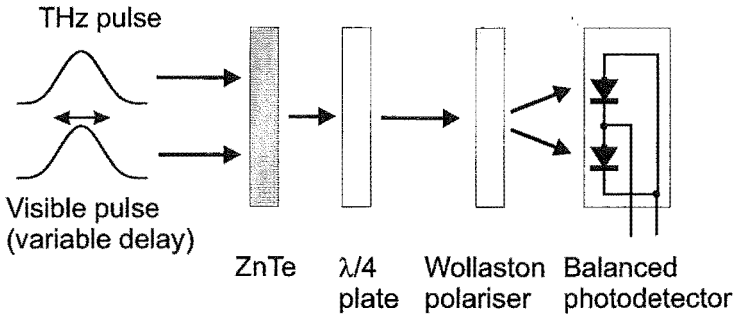


Figure 1.6 Setup for electro-optic sampling of a Terahertz pulse train

Since these photo-conductive switch and optical rectification generation and detection techniques require an ultrafast laser, their large size restricts them to operation in the laboratory. There is a worldwide effort to develop much smaller sources, and detectors, so that portable terahertz systems may be enabled. Various approaches are under investigation, too many to be covered here. The interested reader is referred to references [40–44] for sources, and references [42, 45–47] for detectors. The overall impression is that given the number of approaches that have been successfully demonstrated, and is now conceivable that the goal of relatively inexpensive, compact sources could be achieved, paving the way for a wider range of terahertz frequency applications.

1.2.3 Present and future applications

A wide range of applications for terahertz radiation have been explored. Spectroscopic applications for astronomy are the most prevalent, or at least, the easiest to predict with certainty. Given that half the luminosity and 98% of the photons emitted since the Big Bang fall into the sub-millimetre and far-infrared, a high level of interest from astronomers is hardly surprising [48]. There are in excess of a dozen satellite programmes either underway, or about to begin, variously investigating interstellar and intra-galactic phenomena, earth's atmosphere, the formation of the solar system, the composition of the Wirtanen comet, and the possibility of extra-terrestrial life [6]. In addition, surface-based (landers) or orbital probes of the atmospheres on Mars, Venus, Jupiter and the moons Europa and Titan have all been proposed [6]. Since it is necessary in these applications to distinguish between crowded spectral lines, highly sensitive detectors are needed. Their operation could conceivably be bolstered by some kind of filtering, especially in the spectral range 300 – 2500 GHz where many of the key species, such as the carbon ion C^+ (1.9 THz) that indicates star forming regions, have their first rotational or vibrational line emissions in the sub-millimetre region of the terahertz spectrum [12]. However, obtaining space qualification for a device is an additional hurdle that must be considered, and suggests that initial applications of new

components may be found sooner in terrestrial application.

On earth, terahertz technology has been extensively applied to plasma fusion diagnostics [49] and gas spectroscopy [50]. Magnetically-confined fusion plasmas are of interest to those developing thermonuclear fusion reactors, in particular the prediction of the onset of plasma instabilities and the resulting non-linearities [6]. Sub-millimetre wave spectroscopy is more challenging than microwave spectroscopy, but is attractive for the analysis of lighter molecules. Greater sensitivity can be achieved at the higher sub-millimetre wave frequencies because the spectra line strength tends to increases with the second, or even third, power of frequency. Aside from cataloguing the specific spectra of molecules, there are several possible future uses of terahertz spectroscopy. These involve the rapid scanning and identification of gases, enabling the remote detection and identification of noxious plumes [51]. More versatile systems like the FAst Scan Submillimeter Spectroscopic Technique (FASSST) [52] and optical pulse terahertz time domain spectroscopy systems (generally using the photoconductive-switch antenna system of section 1.2.2) [53] could be used to identify diverse spectra such as simple thermal absorption in gases or dangling bonds on the surfaces of solids [6]. With look-up tables, the identification of gases from the measured spectral line frequency is fast (tens of milliseconds), and the technique could be potentially extended to detecting DNA signatures through dielectric resonances arising from phonon absorption [6].For example, it has been shown that the transmission spectra of salmon and herring differ [54].

Imaging applications are also restricted to the laboratory, but it is here that the most accessible examples of the technology have been produced. So far, two commercial systems have been produced: Picometrix [55], and Teraview [5]. There are two basic types of image that may be formed with the 2-D real-time electro-optic-sampling T-ray imagers that are presently favoured: absorption, and time-of-flight. The absorption measurements can be processed in conjunction with the known thickness to map the distribution of water content in a leaf [1], air-density in a flame (hot air is less dense, and less absorbing) [2], fat content in meat (meat contains more water, and is therefore more absorbing than fat) [5]. Time of flight measurements can be used in conjunction with the known thickness to map the change in refractive index, or with known refractive index, to map the change in thickness [5]. Measuring the water content, and wrinkles, in skin might become a necessary application if cosmetics companies in the USA are required by law to provide evidence regarding the effectiveness of cosmetics. In the visible, birefringence changes are a marker of tissue damage [56], and polarisation difference imaging is used to reveal the boundaries of skin cancers [57] and identify objects obscured in scattering media such as milky water [58]. It may be possible to extend these techniques to the terahertz regime for deeper imaging [3] or other applications [59].

Due to the security, and frequency re-use, benefits associated with the use of ter-

ahertz radiation (see section 1.2.1), wireless communications applications have been proposed [60]. However, predictions concerning the future possibilities of terahertz wireless communications systems involve an element of speculation. A commercial market, rather than research funding, would be required to drive the development of such a system, since there would be little scientific benefit as compared to other applications. Unfortunately, communications market is unpredictable, as may be evidenced by the unexpectedly high bids in the recent (2000) spectrum auction in the United Kingdom for the 3rd generation mobile phone network (2 GHz carrier frequencies). Coincidentally, the total amount of the successful bids for the five available licenses was £22.5 billion [61], which must form a significant component of the cost of entry into such a market should the license options be realised. Since this is the price that providers are prepared to pay for a five-fold increase in bandwidth over that provided by the current 2.5G systems, it suggests the following. If the trend of increasing bandwidth requirements continues, and if the terahertz region remains license-free (possible due to the short propagation distances involved) or licenses are inexpensive compared to other bands (possible if the wide bandwidth available exceeds demand), then terahertz frequency wireless communications systems may become a reality.

1.3 MOTIVATION

This section outlines the motivation for investigating the particular devices studied in this thesis. Two types of device were studied - a variable retarder (chapters three to seven) and a variable filter (chapter eight).

Fixed and variable retarders are well known in the visible, and are of great use in polarisation control. Many retarders for the visible use materials, such as the naturally-occurring uniaxial crystals calcite and quartz, that are unsuitable for use at terahertz frequencies where they exhibit poor optical properties, such as high dispersion and high loss. The use of artificial dielectric techniques allows the necessary birefringence to be developed in ordinarily isotropic materials, such as silicon, that exhibit low dispersion and low loss across the frequency range of interest. Fixed retarders using artificial dielectric materials are already known [62], and have been demonstrated [63,64], but a variable retarder using artificial dielectrics has yet to be demonstrated [65]. One such device has been proposed [65], but the design has yet to be experimentally validated. Should polarisation aspects be as important at terahertz frequencies as they are elsewhere, such as in communications [66], (and there is no reason not to think so), then demonstrating a variable retarder would be a useful step in the world-wide journey to exploiting terahertz frequencies more fully.

Frequency selective filters are another well known type of device, and not just in the visible, but also in electronics, from DC to millimetre waves. There are fewer difficulties in implementing optical filters at terahertz frequencies, since they do not rely on

the specialised anisotropic material properties required for retarders. Therefore, it may be easier to directly transfer optical filters to terahertz frequencies. However, recent developments in the field of photonic crystals have raised the possibility of creating compact, robust, and inexpensive tunable filters, that would be ideal for deployment at terahertz frequencies. This is worth investigating since it could be important to developers of portable terahertz equipment such as spectrometers. Few, if any, such systems are directly under development now, since there is still a great deal of research to be completed on vital components such as sources and detectors of terahertz. However, there is no reason not to begin development of the other components such systems would require. It is also possible that the development of suitable tunable filters could complement the development of sources and detectors, and bring terahertz systems "closer to market". For example, a simple broad-band source could be paired with a tunable filter to create a tunable narrow-band source for use in a spectrometer or a multi-channel communications system.

1.4 AIM AND SCOPE

It is the aim of this study to demonstrate at 100GHz, a quasi-optical variable polarisation compensator and a quasi-optical variable filter that may be used at terahertz frequencies, using materials, fabrication techniques and physical phenomena that will enable the devices to be scaled to operate anywhere in the terahertz region. The 100GHz design frequency was chosen because it is within the measurement range of a state-of-the-art W-band (70 - 110 GHz) free space measurement system to which the author has had access. Although 100GHz is at the lower end of the terahertz region, the measurements are still challenging. A consequence of choosing this frequency range for the prototyping is that the micromachining fabrication is complicated by the need for large structures (for 100GHz, grating periods $\geq$ 0.5mm). However, the advantage is that, if devices can be successfully micromachined with such large structures, then the fabrication process may be easily scaled to produce the smaller featured devices required for operation elsewhere in the terahertz region. Investigating the fabrication-process-dependent small-feature-size limit of these devices is outside the scope of this study, and is more in line with investigations conducted by others in the research group [67, 68].

The chosen device functionalities are fundamental in other frequency ranges, and are likely to be suitable for use with any of the terahertz applications outlined in the last section. It is not the aim of this study to demonstrate these devices within the context of a specific application. The variability and compactness of these devices is unprecedented in the targeted frequency range, and a successful demonstration of the principles involved (this study) must necessarily precede the development of any application that is well placed to take advantage of them (possible future studies).

Therefore it is desired to demonstrate the fundamental behaviour, in good agreement with the developed theory. Well explained, this would provide the necessary starting point for anyone wishing to adapt the designs for use in a specific application, as and when such an opportunity arose.

It should be noted that the design of a variable device is more involved than that of a fixed device. Variability, in this case mechanically actuated, brings with it advantages and disadvantages. Advantageously, one device can do the work of several. For example, a wireless-local area network (W-LAN) component manufacturer could produce one design for the transceiver units, and fix the channel just before shipping by adjusting a grub screw controlling the filter. Or, the same filter could be used to make the transceiver frequency-agile, that is, tunable 'on the fly'. And, depending on the mechanical actuator, any device state could be held indefinitely without power, unlike electrically controlled devices that must necessarily consume energy to maintain any non-rest state. The device may be carefully tested in each state, at rest, using well-behaved actuators such as micrometers. Disadvantages are that, mechanically variable devices are subject to mechanical shock, vibration and resonance, a finite mass must be moved to change states, and additional positional feedback mechanisms may be required for closed loop control. However, these problems are well understood in the field of mechanical engineering, and fall outside the scope of this study.

1.5 OUTLINE

This section provides an outline of the contents and organisation of the remainder of the thesis.

Generally, the original work in this thesis is presented in the order in which it was performed, as shown in Figure 1.7, with the order in which work was performed in Figure 1.7(a), and the chapter location of each piece of work in Figure 1.7(b). Each of the chapters, starting with the background chapter (not included in Figure 1.7), will now be outlined in turn.

The original work performed in this thesis progressed, in order, around a loop consisting of fabrication, measurement, and simulation, as shown in Fig. 1.7(a). The initial retarder (VADR) design proposal did not originate in this thesis, and therefore is separated from the other tasks in the diagram. Work began with the fabrication of a variable retarder (VADR) (i), followed by measurement (ii), simulation (iii), and simulation of improved designs (iv). Then, having completed the retarder work, simulation of the tunable filter was conducted (v), followed by fabrication (vi) and measurement (vii) of a prototype. The chapter location of the write-up of these tasks is summarised in Fig. 1.7(b), and described in the remainder of this section.

Chapter two provides an extensive background to the variable retarder and the tunable filter, and contains three main sections. The first two sections relate only to

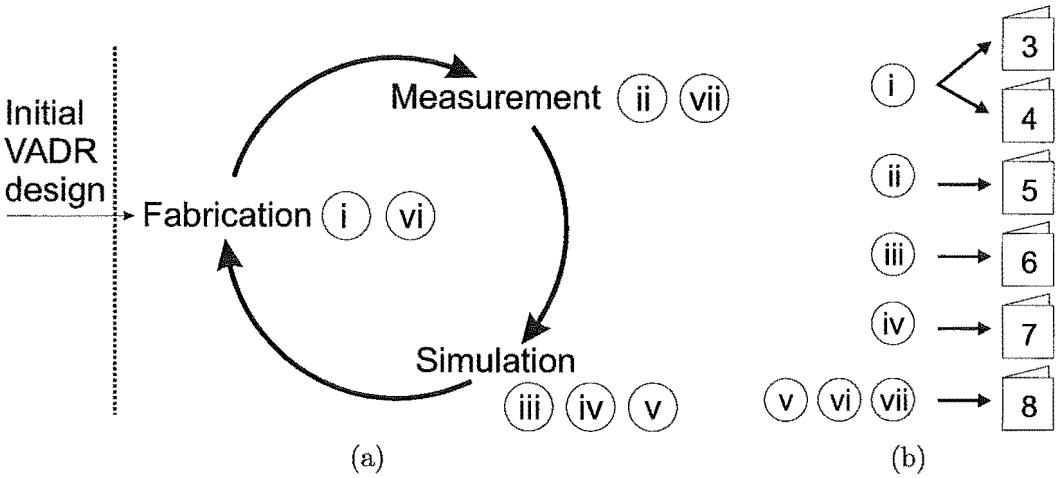


Figure 1.7 Organisation of the original work in this thesis. (a) order work was performed, (b) chapter location.

the variable retarder, whilst the third relates only to the tunable filter. Therefore, the background to each device may be read independently of each other. The first section introduces the polarisation of electromagnetic waves, and sets out the convention for transverse electric (TE) and transverse magnetic (TM) linear polarisations adopted in this thesis. At the end of the first section, uniaxial-crystal retarders for the visible region are outlined, since they are related to the design of the variable artificial dielectric retarder. In the second main section, artificial dielectrics are introduced. It is explained that artificial dielectrics may be used to create birefringence in non-birefringent materials such as silicon, and provides mathematical tools to analyse the birefringence of 1-D subwavelength gratings. The application of 1-D subwavelength gratings to artificial dielectric retarders is discussed, along with the proposal for the variable artificial dielectric retarder studied in this thesis.

The third main section of chapter two moves right away from polarisation, and introduces photonic band gap structures as a means for creating tunable filters at terahertz frequencies. The history of the photonic band gaps is briefly covered. A 1-D photonic crystal is then analysed with photonic band gap analysis techniques, but also with existing, and well known, optical methods, in order to emphasise that the photonic band gap phenomena is completely predicted by Maxwell's equations and that therefore any suitable solution to Maxwell's equations is capable of predicting the presence of photonic band gap along a small number of propagation directions. The effect of defects and conductive substrates are discussed, as are several schemes for creating tunable filters with photonic crystals. The photonic crystals section of the chapter ends with a section covering some competing, non-photonic-band-gap tunable filter designs. The chapter is concluded in the usual fashion.

Chapter three explains the silicon micromachining fabrication processes used in the

construction of the VADR polarisation compensator. Since this chapter is also intended to serve as a reference for use by others in the University of Canterbury laboratory, the device-specific fabrication details are left until the next chapter, where they may also be more clearly described.

Chapter four presents the process for, and results of, the fabrication of the silicon plates for the first device, the VADR polarisation compensator. The mask designs, and simulated results of the mask alignment pattern pre-etch are also included.

Chapter five details the W-band (70 - 110 GHz) free-space vector network analyser, the experimental procedure, and the measurement results for the polarisation compensation device.

Chapter six deals with the simulation of the polarisation compensator. A range of candidate simulation techniques are reviewed, and a detailed comparison is performed of the three techniques used in chapters 6 - 8. Results are presented of simulating the polarisation compensator. An estimate of the free-carrier absorption loss in the low-resistivity substrate is made by fitting to experiments of a plain silicon etalon, and the results are compared to a suitable theoretical model.

Chapter seven presents detailed simulations of improvements to the polarisation compensator, including extensive investigations of the effect of device dimensions on performance. Possible applications of the device are suggested.

Chapter eight introduces the variable photonic band gap filter that was inspired by one of the improved VADR devices. A diverse range of operating modes are presented, along with bandgap maps summaries of extensive simulations into the effect of the fabricated dimensions on the device performance. An elegant device is devised that can be manufactured using conventional metal milling techniques. This is fabricated and measured, showing excellent agreement with the predicted performance. Fabrication of a further device, yet to be tested, is also presented.

Chapter nine is the last in this thesis, and presents the conclusions. It also outlines plans for further work, some of which is currently being undertaken by the author.

Chapter 2

BACKGROUND

2.1 INTRODUCTION

This chapter provides background information relating to both the variable retarder and the tunable filter devices that are studied in this thesis. While both devices are constructed from structured materials (various types of grating, in particular), the principles involved for each, are different. Therefore, the background to each device is presented separately.

Methods of polarisation control that are implemented in the visible region, such as uniaxial crystal waveplates and compensators, can not be successfully scaled to operate at terahertz frequencies, due to their poor optical properties in this regime. For example, the uniaxial crystal calcite exhibits strong dispersion and is opaque in certain ranges of the terahertz band. Instead, the technique of artificial dielectrics may be used to create form-birefringent structures in the surfaces of materials having low disperson and low loss, such as silicon. Fixed-retardance waveplates have been constructed for use in the visible region using artificial dielectrics, but no variable artificial dielectric retarders of any design have been demonstrated prior to this study.

In other applications, it is important to be able to discriminate between different terahertz frequencies. Photonic band gap (PBG) structures offer tantalising advantages for the creation of frequency selective filters, such as strong rejection in a compact device. PBGs arise as a result of nothing more exotic than the destructive interference of multiply-reflected waves within a periodic structure, preventing transmission for a (broad) band of frequencies. Therefore the presence of a PBG can be predicted by a variety of solutions to Maxwell's equations, each having its advantages and disadvantages. By fabricating defects into the periodic structure of a photonic crystal it is possible to realise a narrow band pass within the PBG of the crystal. If the nature of the defect may be altered at will after fabrication, then a tunable filter is realised.

2.2 POLARISATION OF ELECTROMAGNETIC WAVES

This section describes the polarisation[1] conventions for electromagnetic waves used in this thesis. The polarisation is an important property of a wave because it strongly affects how it interacts with matter. Examples of this will arise in subsequent sections.

2.2.1 Uniform Plane Waves

The equation describing a uniform plane wave may be derived from Maxwell's equation, following the method in Appendix A.1, or it may be found in any one of a number of electromagnetic texts, for example [69]. In a right-handed Cartesian co-ordinate system, where the region is homogeneous, linear, isotropic and source-free, the electric field of a monochromatic plane wave, travelling in the positive z direction, with the electric field vector $\mathbf{E}$ aligned to the x axis, may be written

$$\mathbf{E} = E_x \sin(wt - kz)\hat{\mathbf{x}}, \tag{2.1}$$

where w is the angular frequency, k is the propagation constant, t is time, the bold face denotes a vector quantity, and the hat $\hat{}$ denotes a unit vector.

The plane wave of Equation 2.1 is drawn in Figure 2.1, with the electric field plotted along the z axis in Figure 2.1(a), and the path traced by the tip of the electric field vector in Figure 2.1(b), as seen from behind the wave looking in the z direction. This wave is defined to be linearly polarised because the path traced by the tip of the electric field vector is a straight line.

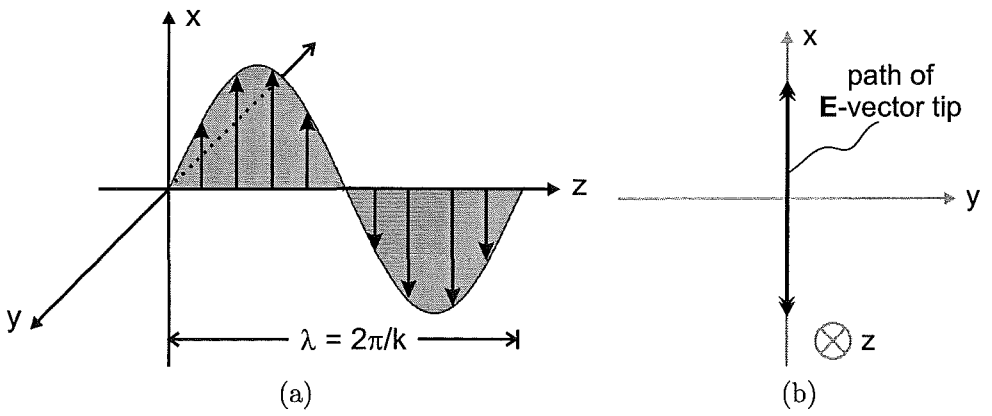


Figure 2.1 A monochromatic uniform plane wave travelling in the positive z direction of a right-handed Cartesian coordinate system: (a) electric field plotted along the z axis, (b) straight-line path traced by the tip of electric field vector.

[1] refers to the classification of electromagnetic waves according to the path traced by the tip of the electric field vector; used in this context extensively throughout the remainder of the thesis, and not to be confused with the dipole moment density of atoms and molecules in solids.

In order to illustrate the three simple polarisation states (linear, circular and elliptical), it is helpful to consider a plane wave with two orthogonal electric field components, E_x and E_y, both travelling in the positive z direction

$$\mathbf{E} = E_x \sin(wt - kz)\hat{\mathbf{x}} + E_y \sin(wt - kz + \phi)\hat{\mathbf{y}}, \tag{2.2}$$

where ϕ is the difference in phase between the two electric field components. The path traced by the tip of the electric field vector may assume either a straight line form similar to Figure 2.1(b), an elliptical form, or a circular form, depending on the relative amplitudes of E_x and E_y, and the phase difference ϕ. Accordingly, the polarisation is defined as being either "linear", "elliptical" or "circular". Linear and circular polarisations are special cases of the more general elliptical polarisation, and are dealt with first.

2.2.2 Linear polarisation

For the sake of comparison with the circular and elliptical polarisations that will be explained in due course, a representative linear polarisation arising from a plane wave with two orthogonal electric field components, E_x and E_y, is shown in Figure 2.2, with the electric fields in Figure 2.2(a), and the path traced by the tip of the resulting electric field vector in Figure 2.2(b). By definition, in a linearly-polarised wave, the two electric field components (E_x and E_y) are always in phase, i.e. $\phi = 0$ or π, since this constrains the tip of the resulting electric field vector to follow a linear path. However, the angle of the path with respect to the x and y axes depends on the relative magnitudes of E_x and E_y. If $E_x = 0$ and $E_y \neq 0$ then the wave is linearly polarised in the y-direction, and vice versa. If E_x and E_y are both non-zero, then the angle to the x axis, (θ in Figure 2.2b), can be found from

$$\theta = \tan^{-1} \frac{E_y}{E_x}. \tag{2.3}$$

2.2.3 TE and TM linear polarisation conventions

In practice, the angle of the electric field for a linear polarisation is normally referenced to fixed direction, such as the axis of one the optical elements. In the analysis of gratings, such as those used in the devices studied in this work, it is the usual practice to define as the reference direction, a grating vector ($\mathbf{K}$), that is orthogonal to the grating. This is important because the properties of the grating strongly depend on the angle of the electric field.

In the measurements and simulations of the variable artificial dielectric retarder (VADR) device, presented in chapters five and six, it is necessary to distinguish be-

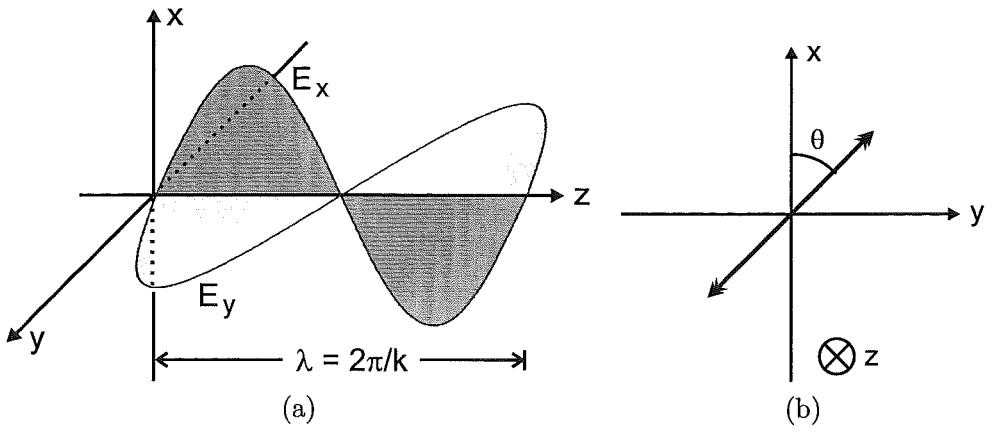


Figure 2.2 Linearly-polarised plane-wave with two orthogonal electric field components, E_x and E_y. The electric fields are plotted along the z axis in (a) and the wavelength λ is indicated. The path traced by the tip of the resultant electric field vector is shown in (b), and the angle θ of the path to the x axis is indicated. The condition of a linearly-polarised wave is that E_x and E_y are in phase, i.e. the phase difference $\phi = 0°$

tween normally-incident, linearly-polarised plane waves that have their electric field aligned either parallel to perpendicular to the grating vector. The adopted convention is depicted in Figure 2.3, for an arbitrary grating. For normal incidence, the transverse-electric (TE) polarisation is a linear polarisation with the electric field perpendicular to the grating vector, as shown in Figure 2.3(a), and the transverse-magnetic (TM) polarisation is a linear polarisation with the electric field parallel to the grating vector, as shown in Figure 2.3(b). The magnitude of a TE-polarised electric field may be denoted by E_{TE} while the magnitude of a TM-polarised electric field may be denoted by E_{TM}. Note that these conventions should not be confused with other conventions, such as those for naming the electromagnetic modes of guided waves. In some texts, the designations P and S are used [70]. These are equivalent to the TE and TM polarisations, respectively.

2.2.4 Circular polarisation

Circular polarisation results if the x and y components are exactly equal in amplitude ($E_x = E_y$), and are in phase quadrature, i.e. they are out of phase by a quarter-wave ($\phi = \pm\pi/2$). Such a wave is shown in Figure 2.4, with the electric fields plotted along the z axis in Figure 2.4(a) and the path traced by the tip of the resulting electric field vector in Figure 2.4(b). The wave may be described by

$$\mathbf{E} = E_x \sin(\omega t - kz)\hat{\mathbf{x}} + E_y \cos(\omega t - kz)\hat{\mathbf{y}}, \tag{2.4}$$

where $E_x = E_y$. As the wave propagates, the electric field vector rotates either clockwise or anti-clockwise. By convention, the wave described by Equation 2.4 is left-

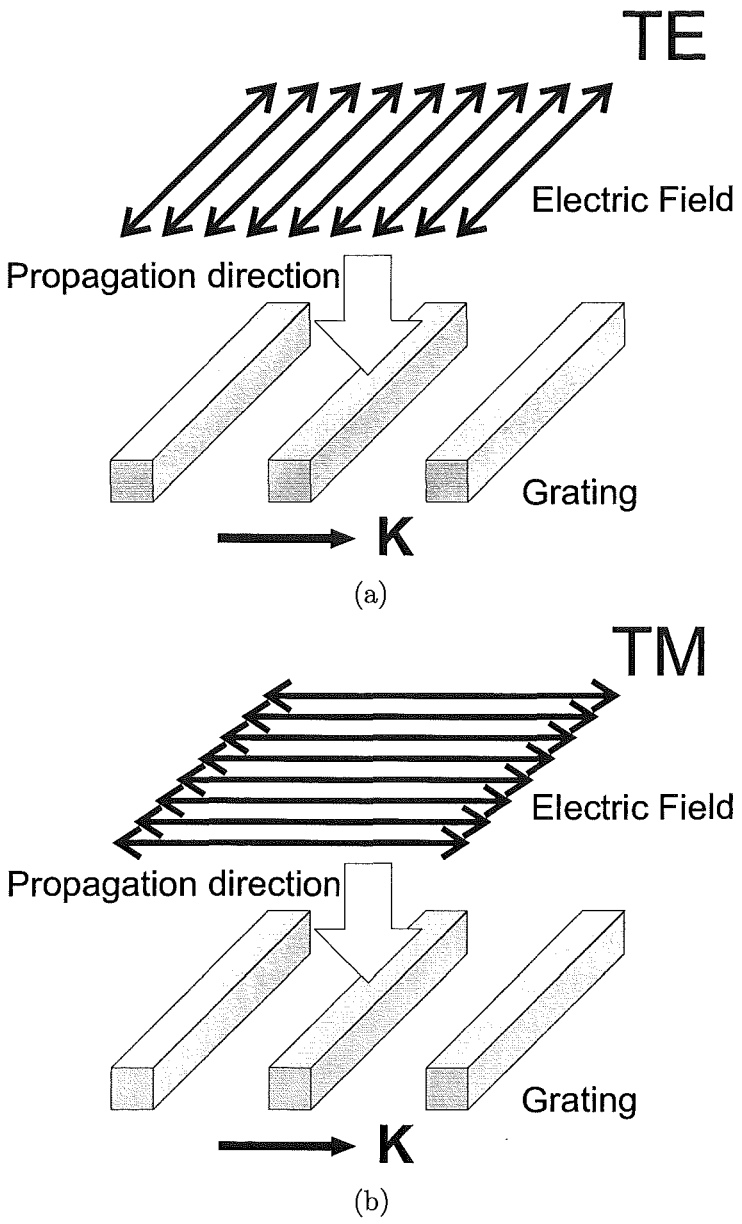


Figure 2.3 The transverse electric (TE) and transverse (TM) polarisations are defined as normally-incident linearly-polarised electromagnetic waves having the following electric orientation with respect to the grating vector **K**: (a) TE - electric field perpendicular to **K**, (b) TM - electric field parallel to **K**.

handed. Standing behind the wave, looking in the positive z direction, the electric field vector appears to rotate anti-clockwise, as shown in Figure 2.4(b).

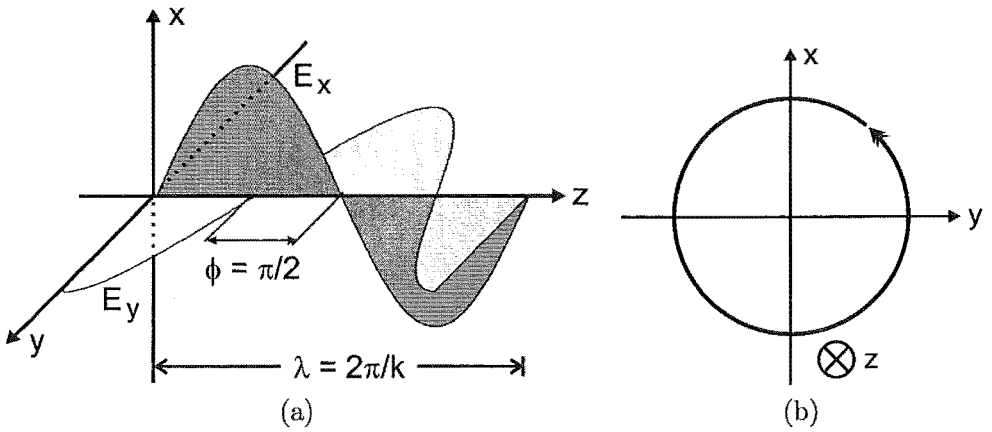


Figure 2.4 Circularly-polarised plane-wave with two orthogonal electric field components, E_x and E_y. The electric fields are plotted along the z axis in (a) and the wavelength λ is indicated. The path traced by the tip of the resultant electric field vector is shown in (b). The condition of a circularly-polarised wave is that E_x and E_y are equal in magnitude and are in phase quadrature, i.e. $E_x = E_y$, and the phase difference $\phi = \pi/2$.

2.2.5 Elliptical Polarisation

In all cases other than linear ($\phi = 0, \pi$) or circular ($\phi = \pm\pi/2, E_x = E_y$), an elliptical polarisation results. Elliptically polarised waves have the same convention for handedness as circular waves, but to characterise them completely it is also necessary to determine both the inverse tangent of the ratio of the semi-major and semi-minor axes (the ellipticity, χ) and the angle of the semi-major to the reference direction (the tilt angle, θ). An elliptical wave is drawn in Figure 2.5 for the case $E_x = E_y$, and $\phi = \pi/4$, with the electric fields plotted along the z axis in Figure 2.5(a) and the path traced by the tip of the resulting electric field vector in Figure 2.5(b). The length of semi-major axis is A, while the length of the semi-minor axis is B. Thus, the ellipticity χ may defined as

$$\chi = \tan^{-1}\left(\frac{B}{A}\right). \tag{2.5}$$

The tilt angle θ is obtained from Equation 2.3, in the same manner as for the linear polarisations, except that for an elliptical polarisation it gives the angle between the x-axis and the semi-major axis.

It should be noted that the ellipticity of a plane-wave's polarisation may be varied from linear to elliptical, and if $E_x = E_y$, to circular, simply by changing the phase difference ϕ between the two orthogonal electric field components, E_x and E_y.

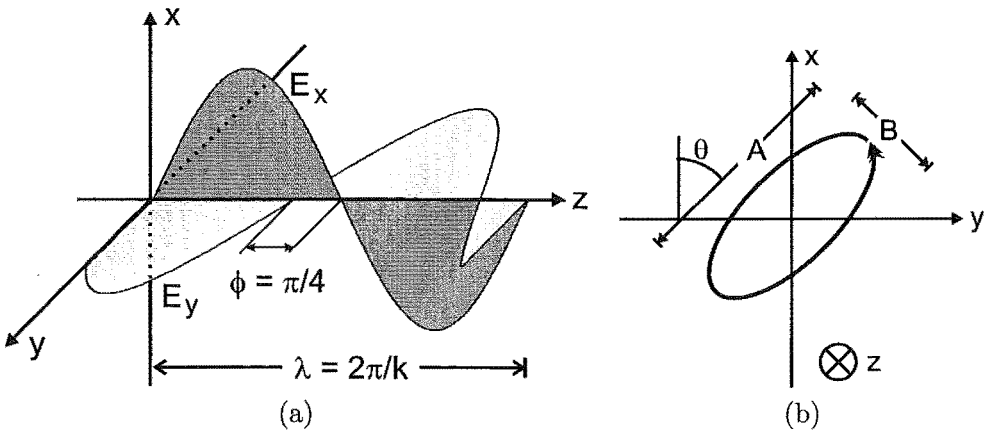


Figure 2.5 Elliptically-polarised plane-wave with two orthogonal electric field components, E_x and E_y. The electric fields are plotted along the z axis in (a) and the wavelength λ is indicated. The path traced by the tip of the resultant electric field vector is shown in (b), and the angle θ of the path to the x axis is indicated.

2.3 NATURALLY-BIREFRINGENT CRYSTAL RETARDERS

A variety of devices have been implemented for the control of polarisation in the visible and infrared, including beam splitters, polarisers, retarders, and rotators [71,72]. It is not intended to review all these devices; instead, this section is restricted to a discussion of several relevant designs of birefringent retarder that use uniaxial crystals. It is in part from these designs that the VADR device (studied in this thesis) has evolved. While the frequency ranges, and implementations, are different, the underlying principles are the same. Should the reader desire further information on devices for polarisation control, the texts of Clarke and Grainger [71], Hecht and Zajac [73] or Born and Wolf [74] may be recommended.

The purpose of a retarder is to alter the polarisation state by introducing an additional *relative* phase delay Γ between the x and y polarisations such that the phase difference after the retarder is

$$\phi = \phi_0 + \Gamma \tag{2.6}$$

where ϕ_0 is the phase difference in the incident beam. For example, it might be desired to change a linear polarisation having tilt angle $\theta = \pi/4$ into a circular polarisation, as shown in Figure 2.6. As in the previous section, a plane wave having two orthogonal components is considered

$$\mathbf{E} = E_x \sin(wt - kz)\hat{\mathbf{x}} + E_y \sin(wt - kz)\hat{\mathbf{y}}. \tag{2.7}$$

In order to change the polarisation, a retarder introduces a *relative* phase delay Γ,

giving the output wave as

$$\mathbf{E} = E_x \sin(wt - kz)\hat{\mathbf{x}} + E_y \sin(wt - kz + \Gamma)\hat{\mathbf{y}}. \tag{2.8}$$

The amount of *relative* phase delay Γ introduced by the retarder is a key property, and is known as the "retardance". In the remainder of this thesis, retardance is denoted by the symbol Γ. Continuing with the example, if the plane wave of Equation 2.7 is incident on a retarder having a retardance of $\Gamma = \pi/2$, then the incident linear polarisation will be converted into circular polarisation. In this example, the output circular polarisation is left handed.

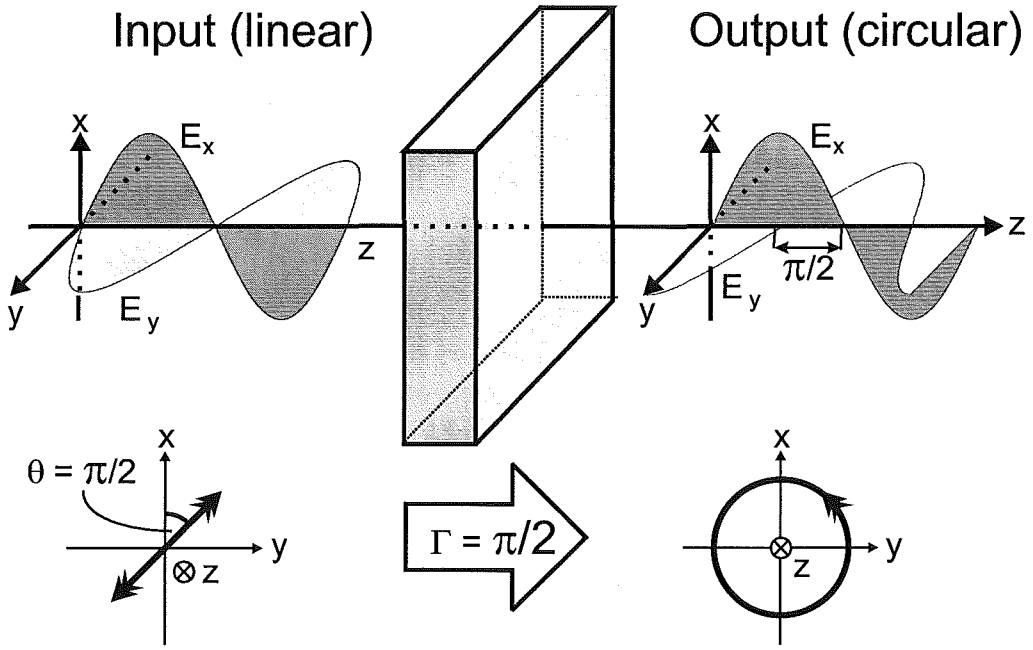


Figure 2.6 An example of a quarter-wave retarder changing a linear polarisation (input, left hand side) into circular (output, right hand side). In this case, the E_x component is delayed by $\Gamma = \pi/2$ upon transmission through the retarder. At the bottom are plots of the path traced by the tip of the electric field vector, looking in the direction of propagation.

Waveplates (fixed retarders) usually have a retardance of either $\Gamma = \pi/2$ (quarter-wave) as in the previous example, or $\Gamma = \pi$ (half-wave), since with an arrangement of three rotatable waveplates ($\Gamma_1 = \frac{\pi}{2}, \Gamma_2 = \pi, \Gamma_3 = \frac{\pi}{2}$) any arbitrary polarisation transformation can be achieved [75]. The actions of quarter- and half-wave waveplates may be summarised by the polarisation transformations listed in Table 2.1. The tilt angle θ is the same as that defined in Equation 2.3 for linear and elliptical polarisations, as defined in sections 2.2.2 and 2.2.5. No tilt angle is defined for circular polarisations because the path traced by the tip of the electric field vector is rotationally invariant. The terms "Right" and "Left" refer to the handedness of the elliptical and circular polarisations, where for a right handed polarisation the tip of the electric field vector rotates in a clockwise direction when seen from behind, and for a left handed polarisation,

anti-clockwise (see section 2.2.4).

Table 2.1 Polarisation transformations of quarter- and half-wave waveplates (After [76] Table 1).

Quarter-wave ($\Gamma = \pi/2$)	
Input	Output
Linear, $\theta = 45°, \phi = 0$	Right Circular,$\phi = \pi/2$
Linear, $\theta = -45°, \phi = -\pi$	Left Circular $\phi = -\pi/2$
Right Circular $\phi = \pi/2$	Linear, $\theta = -45°, \phi = \pi$
Left Circular $\phi = -\pi/2$	Linear, $\theta = 45°, \phi = 0$
Linear, any $\theta \neq 45°, \phi = \phi_0$	Elliptical, $\phi = \phi_0 + \pi/2$
Half-wave ($\Gamma = \pi$)	
Input	Output
Linear, angle $\theta, \phi = 0$	Linear, angle $-\theta, \phi = \pi$
Left circular,$\phi = -\pi/2$	Right circular, $\phi = \pi/2$
Right circular,$\phi = \pi/2$	Left circular,$\phi = -\pi/2$

Birefringent waveplates (fixed retarders) for visible wavelengths are traditionally made from a carefully cut and polished slab of uniaxial crystal, such as quartz or calcite. The term "uniaxial" is used to describe these crystals because because the optical properties along one axis differ from those along the other two. This is illustrated in the drawing of Figure 2.7 that represents a slab of calcite that has been cut and polished for use as a waveplate. A wave propagating in the $(\hat{e} \times \hat{o})$ direction (the direction orthogonal to the extraordinary and ordinary axes), with the electric field oriented along $\hat{e}$, will experience a material having refractive index n_e. If the electric field is instead oriented along $\hat{o}$, the wave will experience a material having refractive index n_o. The refractive index n_o that is specified in the $(\hat{e} \times \hat{o})$ direction does not affect either of these waves, and would only affect a wave propagating in $\hat{e}$ or $\hat{o}$ with its electric field oriented in the $(\hat{e} \times \hat{o})$ direction. For calcite, since $n_o < n_e$, a wave with its electric field oriented along $\hat{o}$ will propagate faster than one with its electric field oriented along $\hat{e}$. Therefore, the E_e component is delayed by the relative phase delay (retardance) Γ (alternatively, it may be conceived that the E_o component is advanced by Γ with respect to the E_e component).

The unique axis is called either the "optical" or "extraordinary" axis, and has its refractive index denoted by n_e, while the other two axes are called the "ordinary" axes and have their refractive index denoted by n_o. The axis (or axes) with the smallest index is known as the fast axis (or axes), and when the extraordinary axis is fast, the crystal is known as a positive uniaxial crystal. By this definition, quartz is a positive uniaxial crystal, while calcite is a negative uniaxial crystal.

To a first approximation, the retardance (Γ) of a uniaxial crystal slab, having thickness d, is equal to the difference in the effective path length in the direction of

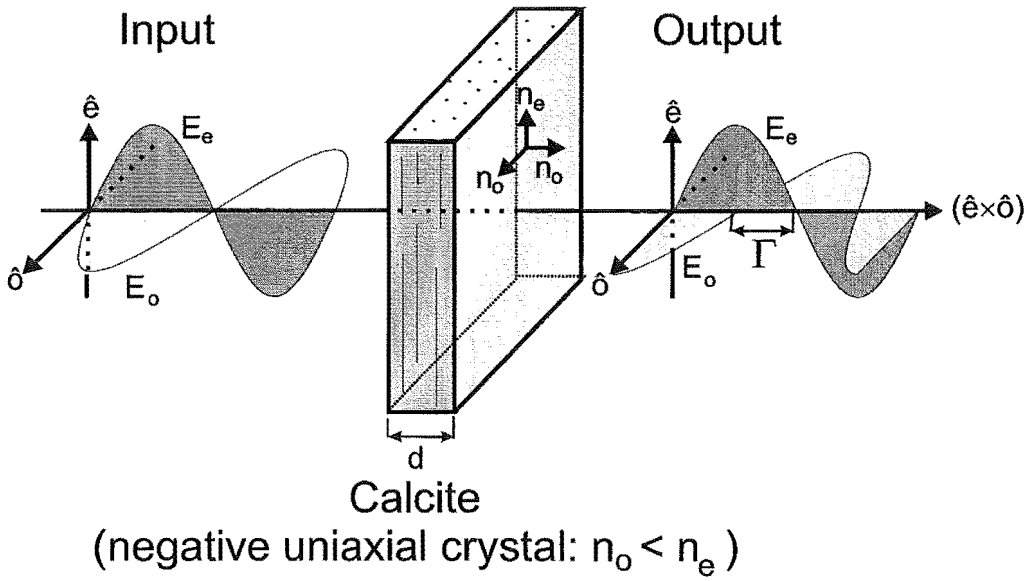


Figure 2.7 A calcite waveplate, showing the ordinary $\hat{o}$ and extraordinary $\hat{e}$ directions in the crystal, and the associated refractive indices n_e and n_o. Lines and dots are used to indicate the direction of the extraordinary axis - these are a guide to the eye and are not seen on an actual crystal.

propagation

$$\Gamma = \frac{2\pi d}{\lambda_0}(n_e - n_o). \tag{2.9}$$

This approach is sometimes known as the "amplitude transmittance approach". Unfortunately, this simple approach is inaccurate because it does not take account of multiple reflections from the air-dielectric interfaces at the front and back surfaces of crystal slab. The error is correspondingly greater when the index discontinuities at the front and rear surfaces are large, because the reflections are stronger. For uniaxial crystal retarders at visible wavelengths, the error may be as much as 5%, which is undesirable for accurate polarimetry [77].

A more accurate model includes these reflections [78], but is not presented here because it does not account for the loss in the crystal. However, the model does lead to one important conclusion that is not apparent from Equation 2.9. Since the transmission coefficients are dependent on the reflectivities at the front and rear surfaces, they differ for the ordinary and extraordinary rays because the crystal is birefringent. Consequently, the slab must be rotated to offset the slight polarisation rotation that this introduces. Alternatively, the asymmetry of the transmission coefficients may be reduced by coating the front and rear surfaces with an anti-reflection quarter-wave dielectric coating [69]. The refractive index of a quarter-wave coating n_{qc} is given by the geometric mean of the refractive indices of the materials it interposes, n_i for the

incident medium and n_s for the substrate,

$$n_{qc} = \sqrt{n_i n_s} \tag{2.10}$$

and its thickness is given by

$$d_{qc} = \frac{\lambda_{n_{qc}}}{4} = \frac{\lambda_0}{4n_{qc}}. \tag{2.11}$$

Since n_s is anisotropic (due to the birefringence of the crystal), but n_{qc} is not, reflections can not be completely eliminated. If a representative value of n_s is chosen between n_e and n_o, the residual reflections will be small [71].

For naturally birefringent crystals such as calcite, which has one of the highest birefringences exhibited by naturally occurring crystals, it is impractical to cut the waveplate thin enough to obtain zero-order operation (total retardance less than 2π) at visible wavelengths since such thin plates would be prone to breakage. For mechanical strength, the plates are normally cut thicker, to provide a retardance of $N2\pi + \Gamma$, where N is equal to the waveplate's order. At the design wavelength, higher order (thicker) waveplates still have the same effective retardance Γ, but for other wavelengths, the error is greater than for the thin, zero order plate.

It is possible to create a zeroth-order waveplate when a variable design is employed, as will be explained in the rest of this section. Variable devices also have other advantages. Firstly, if the illumination is always monochromatic, but not always of the same wavelength, then a single compensator obviates the need to keep many waveplates on hand, since the former may be tuned to exactly the desired retardance over a range of wavelengths, whereas a different waveplate is often required if the wavelength is altered. Secondly, compensators may be dynamically tuned via a feedback loop to continually adjust for deviations from a desired polarisation state. Two compensator (variable retarder) designs for the visible are now described.

The Babinet compensator, shown in Figure 2.8, comprises two wedges of uniaxial crystal, cut so that the extraordinary axis in one is perpendicular to the extraordinary axis in the other. The wedges are thin, having an angle of about 2.5°. One wedge is fixed in position, while the position of the other is controlled by a micrometer as indicated in Figure 2.8. Again, lines and dots are used to indicate the direction of the extraordinary axis. A wave that propagates vertically down through the device at some point will traverse a distance d_1 through the first wedge and d_2 through the second wedge. In the first crystal, neglecting reflections, the retardance is given by

$$\Gamma_1 = \frac{2\pi d_1(|n_o - n_e|)}{\lambda_0}, \tag{2.12}$$

while in the second wedge the retardance is given by

$$\Gamma_2 = -\frac{2\pi d_2(|n_o - n_e|)}{\lambda_0}. \tag{2.13}$$

The negative sign in the right hand side of Equation 2.13 arises because the extraordinary axes of the two wedges are perpendicular. The retardance in the first wedge becomes offset by a phase advance in the second. If $d_1 = d_2$ there is no net change in the relative phase. Otherwise, the overall retardance, neglecting reflections, is given by

$$\Gamma = \Gamma_1 + \Gamma_2 = \frac{2\pi}{\lambda_0}(d_1 - d_2)(|n_o - n_e|). \tag{2.14}$$

By altering the relative values of d_1 and d_2, either by moving the beam or shifting the wedges with respect to each other, any desired retardance may be obtained. Thus, subtle corrections may be made to deformed polarisation states. Equally usefully, the compensator is not restricted to use at any single wavelength, but provided the illumination is always monochromatic, it may be adjusted to provide a particular retardance across a wider range of wavelengths.

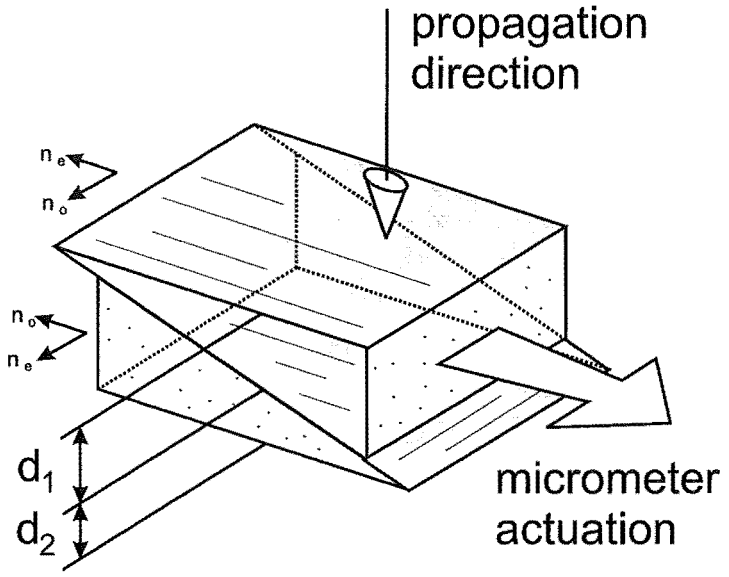


Figure 2.8 A Babinet compensator comprises two wedges of uniaxial crystal with their extraordinary axes oriented perpendicular to each other.

Since the values of d_1 and d_2 depend on where the beam strikes the Babinet compensator, it cannot produce a uniform retardance over its surface. Therefore, if a uniform retardance is required, the incident beam must be narrow. Alternatively, for a broad incident beam, the top wedge may be rotated by 180°, so that the thin end rests on the thin end of the bottom wedge. In this configuration, the beam is slightly deviated because the top and bottom surfaces are not parallel.

If it is desired to obtain uniform retardance across the surface of the compensator, and suffer no beam deviation, then the Babinet-Soleil compensator may be used. The Babinet-Soleil compensator comprises two wedges sharing the same extraordinary axis, and a slab having its extraordinary axis perpendicular to that of the wedges, as shown in Figure 2.9. Where the two wedges overlap, the total thickness of the wedge section d_1 is constant with position, therefore a uniform retardance may be obtained across the whole device. With d_1 defined for the total thickness of the two wedges, and with d_2 the thickness of the slab, the retardance of the Babinet-Soleil compensator is given by Equation 2.14. The top wedge is actuated by a micrometer, and thus the value of d_1 may be altered without affecting the uniformity of the retardation. The disadvantage of a Babinet-Soleil compensator is the expense involved in the precise cutting and polishing of three pieces of crystal.

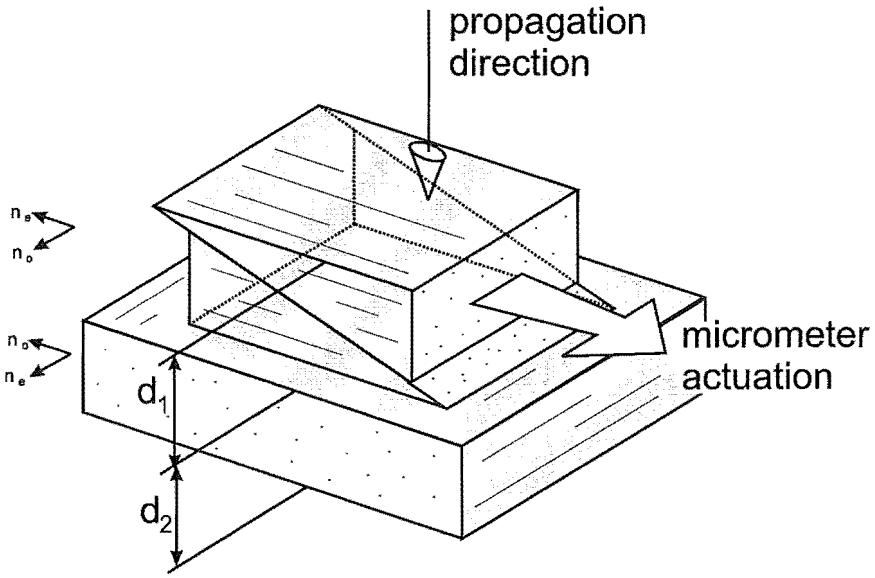


Figure 2.9 A Babinet-Soleil compensator comprises two wedges of uniaxial crystal with their extraordinary axes aligned, mounted on top of a slab of uniaxial crystal that has its extraordinary axis oriented perpendicular to that of the wedges.

Compensators have the property of assuming any value of retardance within a given range, usually $0 \leq \Gamma \leq \pi$. Their transformations may not be neatly summarised in table format; instead it is preferable to use an alternative representation such as Poincaré's sphere [71], pictured in Figure 2.10. Each point on the surface of this imaginary sphere represents a particular polarisation state, specifying the tilt angle θ, the ellipticity χ and the handedness. It is possible to represent the intensity (power) of the wave by the radius r of the sphere, but a constant radius sphere (fixed intensity) may be assumed here to avoid confusion. Often, the three axes are labelled Q, U, V in connection with the Stokes' parameters [71]. The Stokes' parameters may be used to describe the polarisation state of both coherant and incoherant beams, and have the advantage of being able to describe partially polarised beams. For the analysis of

the total effect of a chain of optical elements in series, acting on a coherent, totally polarised beam, then the Jones' calculus is of use. The Stokes' parameters, and the related Jones' calculus are described in Appendix A.2 as they are not used elsewhere in the thesis.

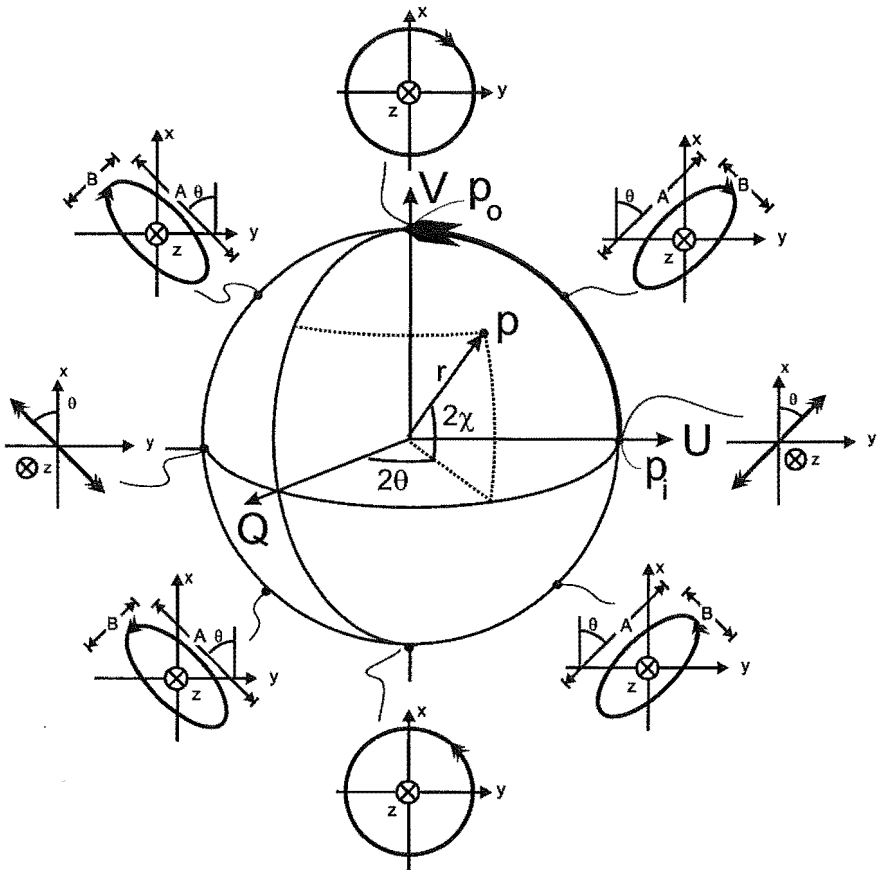


Figure 2.10 The Poincaré sphere representation of polarisation. Any point P on the sphere is uniquely defined by twice the tilt angle θ and twice the ellipticity χ of the polarisation ellipse. The radius r may be used to represent intensity.

In geographical terms, the equator of the Poincaré sphere represents the linear polarisations, with the longitude of the point being twice the tilt angle, or 2θ. The north and south poles represent right and left handed circular polarisations. Points in between are elliptically polarised, with the handedness of the closest pole (right in the northern hemisphere, left in southern hemisphere). The latitude of the point represents the ellipticity, being equal to twice the ellipticity angle defined in Equation 2.5, or 2χ. Several representative polarisation states are indicated in Figure 2.10.

Any possible change in polarisation can be described by a curve on the surface of the sphere that connects the state of polarisation before the change, with the state afterwards. For example, a transition from point P_i to P_o is shown by the block-headed arrow in Figure 2.10. P_i represents the polarisation state of the incident illumina-

tion, while P_o represents the polarisation state of the output of a retarder having a quarter-wave retardance. In this case, the linearly polarised input is converted to a circular output. For other tilt angles ($\theta \neq \pi/4$), it is not possible for the retarder alone to produce a circular output. In these cases, the magnitudes of the two orthogonal components, E_x and E_y, must first be equalised by transmission through a polariser oriented at 45° to the axes of the retarder [73], or by another combination of waveplates.

2.3.1 THz properties of uniaxial crystals

This section is given over to a discussion of the optical properties of uniaxial crystals (calcite, in particular), and the calculated performance of a uniaxial-crystal waveplate, at terahertz frequencies. This section validates the need for alternative retarder implementations at terahertz frequencies.

Calcite is chosen for this example because accurate data is available for its optical properties at terahertz frequencies [79]. This data was obtained from a text that critically evaluated and compiled the published results of a number of researchers. In order to illustrate the difference in the properties between silicon (which is not naturally birefringent but will be used for artificial birefringent devices here) and calcite at terahertz frequencies, the refractive index and attenuation coefficient in the ordinary axis of calcite are plotted in Figures 2.11, and may be compared against similar data plotted for silicon in Figure 1.3. The data for the extraordinary axis are not plotted, since it exhibits similar dispersion and loss. Calcite is strongly absorbing between 2.5 - 10 THz, but it is conceivable that a reasonably transmissive waveplate could be constructed for 0.1 – 2.5 THz (i.e. a subset of the range of frequencies (0.1 – 10 THz) that are of interest in this study).

For the sake of illustration, a specific example is now taken. The properties of a zero-order quarter-wave waveplate, constructed from calcite and having quarter-wave anti-reflection coatings for maximum performance, are quantified using the T-matrix method that is presented in section 6.2.1 (see also Hecht and Zajac [80]). The T-matrix method allows the the transmission coefficient and retardance to be calculated taking account of all reflections, and the loss in the dielectric. The refractive index of the quarter-wave coatings n_{qc} is given by Equation 2.10 where the substrate index is chosen to be the geometric mean of the real part of the refractive indices for the ordinary and extraordinary rays ($n_s = \sqrt{n_o n_e}$), and the coating thickness is given by Equation 2.11. The properties are tabulated in Table 2.2 at frequencies for which accurate data for the properties of calcite were available. The thickness d was arrived at by iteratively calculating the exact retardance Γ of slabs of varying thickness, starting with the thickness obtained by a rearranged form of Equation 2.9. The transmission coefficients are presented in dB for later comparison to the VADR devices studied in this thesis.

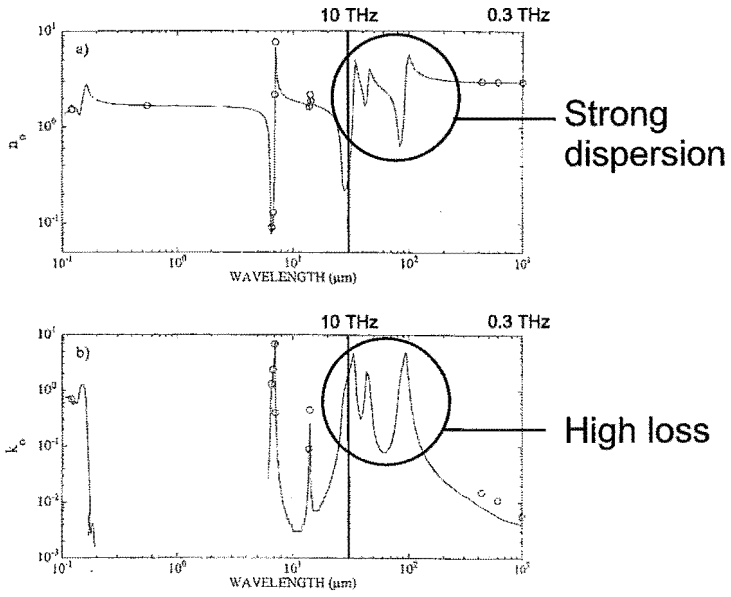


Figure 2.11 (a) Log-log plot of the refractive index of calcite in the ordinary axis (n_o) versus wavelength in micrometres. (b) Log-log plot of the extinction coefficient in the ordinary axis (k_o) versus wavelength in micrometres. Most of the terahertz frequency range of interest is represented (0.3 - 10THz), as indicated.(Reproduced from [79] Fig. 1, p704.)

Table 2.2 Calcite quarter-wave plate for low terahertz frequencies. The symbols are as follows: f is the design frequency, while λ is the corresponding free-space wavelength; n_o, k_o are the refractive index and extinction coefficient for the ordinary ray, while n_e, k_e are for the extraordinary ray; n_{qc} is the refractive index of the quarter-wave antireflection coatings and the slab thickness is d (excluding the coatings), while τ_o and τ_e are the expected insertion losses for the ordinary and extraordinary rays.

f (GHz)	λ (μm)	n_o	k_o	n_e	k_e	n_{qc}	d (mm)	τ_o (dB)	τ_e (dB)
90	3333.3	2.946	0.0	2.883	0.0	1.707	9.6810	-0.8	-2.6
150	2000	2.941	0.002	2.873	0.003	1.705	7.4783	-0.7	-4.5
300	1000	2.950	0.0057	2.862	0.0066	1.705	2.2225	-1.7	-3.5
600	500	2.961	0.007	2.909	0.014	1.713	2.2001	-2.7	-6.5
1050	285.71	3.007	0.014	2.999	0.028	1.733	8.9210	-26.4	-50.2
3000	100	5.677	1.508	1.733	5.206	1.771	0.0031438	-10.4	-12.7
9900	30.303	0.335	2.096	0.339	2.385	0.581	1.8938	<-130	<-130

It may be seen from Table 2.2 that due to the strong dispersion of calcite, the required thickness varies dramatically, from 3μm – $1cm$. With the exception of the extremely thin (but still lossy) 3 THz plate, all the waveplates in the table are over 2mm thick, but still require the surface to be polished to sub-micron flatness, and therefore represent a costly, cumbersome and poorly-performing solution. Moreover, 1 THz is an important experimental frequency [5, 35] yet here the required thickness is a massive 31λ, and the loss figures are completely unacceptable – they exceed the rejection rate of an optical infrared filter [81]. Towards 10 THz, the plate becomes opaque and totally unusable. Therefore, naturally birefringent crystals such as calcite are unsuitable for use at terahertz frequencies, and alternative techniques must be sought so that the entire terahertz frequency range may be covered.

One possible solution is to generate birefringence in silicon, using artificial dielectric techniques. This allows advantage to be taken of the silicon's desirable properties (low dispersion and low loss, as shown in Figure 1.3), properties that are not exhibited by the naturally-occurring birefringent crystals.

2.4 BIREFRINGENT ARTIFICIAL DIELECTRIC MATERIALS

2.4.1 Introduction

As shown above, in the construction of polarisation devices at terahertz frequencies, it becomes necessary to use microfabricated form-birefringent structures in the place of the naturally-occurring uniaxial crystals employed in the visible and infrared, in order to obtain materials with low-loss and low-dispersion. Furthermore, form-birefringent structures are capable of exhibiting stronger birefringence than the natural uniaxial crystals, and thus artificial dielectric devices may be more compact, in addition to being less expensive.

At terahertz frequencies, the required sub-wavelength feature sizes for artificial dielectrics are in the range of $1 - 500$ μm, which is well within the capabilities of micromachining techniques such as the wet etching of silicon. Two examples of micromachined artificial dielectric structures are shown in Figure 2.12, with a non-birefringent anti-reflection surface in Figure 2.12(a) having a 2-D array of square pyramids, and with a birefringent waveplate surface in Figure 2.12(b) having a 1-D array of V-grooves.

Artificial dielectrics are perhaps best known not for their birefringence, but for their role in the reduction of reflections at optical and quasi-optical interfaces [82–84]. Early anti-reflection surfaces in the visible, similar to that shown in Figure 2.12(a), were based on structures found in the eyes of nocturnal insects, having periodic arrays of cones of about 200nm depth and spacing [82]; hence the term "moth-eye effect" is often used to describe such behaviour. It was found by early microwave analogy experiments that these structures increased the transmission of visible light into the insect's eye [85].

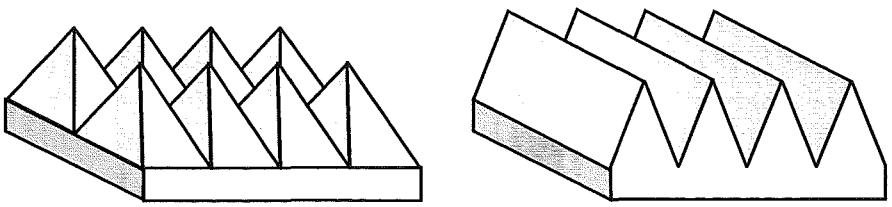


Figure 2.12 Examples of artificial dielectric surfaces having features with subwavelength periodicity. (a) A 2-D array of pyramids makes a useful anti-reflection surface. Surfaces like this mimic those found in nature, for example in the eyes of moths [82]. (b) A 1-D array of V-grooves makes a birefringent surface, such as might be used in a quarter- or half-wave waveplate.

Qualitatively, the reduction in reflection (improved transmission) is achieved because the incident light is unable to resolve the detail of the subwavelength cones on the surface. Instead, it "sees" a homogeneous material with a graded "effective" refractive index, that gradually increases from the refractive index of air ($n_i = 1$) to the value of the substrate n_s (where $n_s > n_i$). The name "artificial dielectric" is given because the graded dielectric constant experienced by the incident light is artificially created, and does not exist independently of the illumination.

Birefringence is created in the surface of Figure 2.12(b) by presenting the incident TE and TM waves with a different grading of the refractive index during the transition between air and substrate. The form of the sub-wavelength structuring "seen" by the TE and TM waves affects the effective refractive indices for these waves, hence the term "form birefringence" is often used.

2.4.2 Effective Medium Theory

The effective index or birefringence must be able to be quantified before artificial dielectric structures may be designed. Two approaches may be taken. Either the electromagnetic properties of the structure may be solved for using rigorous numerical computer codes (examples of this are covered in detail in chapter 6), or an analytical approach may be taken. In this section, the latter approach is taken, and the use of effective medium theory (EMT) of artificial dielectrics is described[2].

Depending on the structure to be analysed, and the particular EMT formulation, the solution is either rigorous or approximate. Even in circumstances when the solution is approximate, EMT may provide useful insight into how the geometry of the artificial dielectric structure affects its performance. Also, EMT is advantageous in situations where it is desired to analyse artificial dielectrics in combination with stacks of homogeneous layers, since it allows the artificial dielectric to be modelled as one or more

[2]The term "effective medium theory" is not exclusively used in connection with artificial dielectrics in the literature, and should not be confused with unrelated concepts.

homogeneous layers that has equivalent properties. As a result, the overall reflection and transmission coefficients may be readily determined with thin-film theory [73] or a transmission line analogy [69], without need to resort to a more involved solution.

The key limitation to the applicability of the EMT approach is that it cannot be used for structures that diffract, because homogeneous layers are not capable of diffraction and therefore do not have equivalent properties. In the next section the "no diffraction" requirement is formalised in terms of the period-wavelength ratio of a 1-D grating. Thereafter, different EMT solutions are presented depending on the scale of the structure and the required accuracy.

The following discussion of artificial dielectrics is limited to gratings that are periodic in only one dimension (1-D), because the artificial dielectrics employed in this thesis are 1-D (see initially the explanation of the VADR device in section 2.5.1). Therefore, an analysis of the two-dimensionally-periodic moth-eye surface of Figure 2.12(a) is not presented. However, an examination of the 1-D theory provides ample coverage of the relevant issues, but without the additional algebra for 2-D artificial dielectrics.

2.4.3 Subwavelength gratings

A periodic structure, such as a grating, may be called an "artificial dielectric" if the transmitted and reflected zeroth diffraction orders are the only propagating orders, and all the higher orders are evanescent [83]. The distinction between artificial dielectric gratings and other gratings is formalised in this section by placing an upper bound on the grating period, in terms of the wavelength, for an artificial dielectric grating.

Figure 2.13 shows a cross-section of an arbitrary 1-D surface-relief grating. The grating vector $\mathbf{K}$ is shown, as is the period Λ, the refractive indices of the substrate, n_s, and the incident medium, n_i, the transmitted and reflected zeroth orders, T_0 and R_0 respectively, and the transmitted and reflected higher orders, T_m, T_{-m} and R_m, R_{-m} respectively, where m is the order.

The grating equation may be used to determine whether a given order propagates or not:

$$n \sin\theta_m - n_i \sin\theta_i = \frac{m\lambda}{\Lambda}, \tag{2.15}$$

where n is the index of the medium that the diffracted order is propagating in (n_s for transmitted orders and n_i for reflected orders), θ_i is the angle of incidence as measured to the normal the grating surface, and θ_m is the angle of the mth diffracted order. An upper bound for the period can be found by setting $m = 1$ (the first evanescent order in an artificial dielectric), $\theta_m = 90°$, and recasting Equation 2.15 as an inequality

$$\frac{\Lambda}{\lambda} < \frac{1}{\max[n_s, n_i] + n_i \sin\theta_{max}}, \tag{2.16}$$

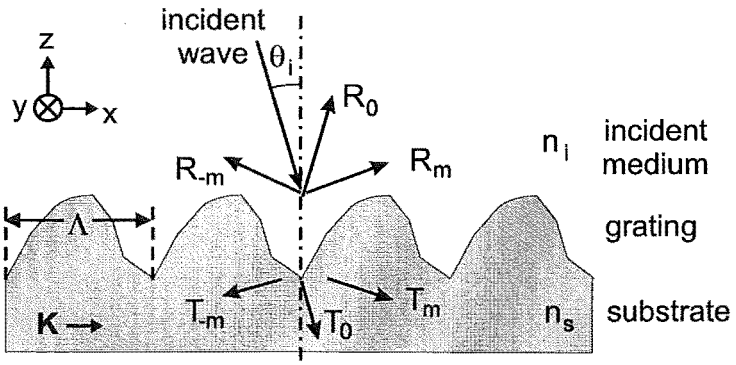


Figure 2.13 Transmitted and reflected diffraction orders for an artificial dielectric grating of arbitrary profile. For a given angle of incidence, θ_i, all diffraction orders except the zeroth order transmitted (T_0), and zeroth order reflected (R_0), are evanescent. For this to occur, the grating period Λ must be smaller than the incident wavelength λ, as specified in Equation 2.16 (After Figure 1, in Reference [83]).

where max[] is equal to the maximum value of its arguments, and θ_{max} is the maximum angle of incidence.

For Equation 2.16, the denominator on the right hand side must be greater than unity in any practical application, enforcing the condition that the period of the grating must be smaller than the incident wavelength. It should be noted that this upper bound on the period is useful only in determining whether a grating may be considered an artificial dielectric. There are additional limitations placed on the region of validity for the EMT analysis, and these will be explained in due course.

2.4.4 Rytov's EMT

In his widely referenced paper of 1956, Rytov studied the properties of stratified electromagnetic media, with the aim of reducing the losses in transmission lines filled with such a medium [86]. The method he used is also applicable to the analysis of artificial dielectric gratings. The stratified media he analysed contained alternate plane layers of two arbitrary homogeneous materials, the first having permittivity ε_1, permeability μ_1 and thickness a, and the second having permittivity ε_2, permeability μ_2 and thickness b, as shown in Figure 2.14. The refractive index n used in the previous section is related to the permittivity ε and permeability μ by

$$n = \sqrt{\frac{\mu}{\mu_0}\frac{\varepsilon}{\varepsilon_0}}, \tag{2.17}$$

where ε_0 and μ_0 are the permittivity and permeability of free space, respectively. The EMT equations are less cluttered with superscripts when presented in terms of the permittivity.

The structure was illuminated with a uniform plane wave, and solved for the exact

fields. The two solutions presented for the plane wave propagating in the z direction[3] are of interest because they correspond to the TE and TM polarisations defined in section 2.2.2.

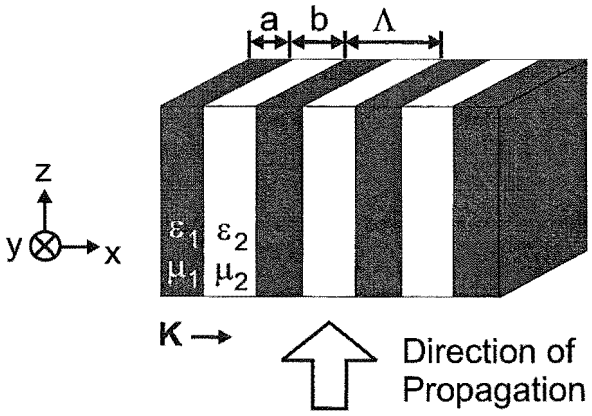


Figure 2.14 The stratified electromagnetic medium analysed in Rytov's paper [86], with alternating layers, the first having permittivity ε_1, permeability μ_1 and thickness a, and the second having permittivity ε_2, permeability μ_2 and thickness b. The period is $\Lambda = a + b$. (Note that the x, y, z directions shown here differ from those in the paper.)

Subsequently, solutions to the average electromagnetic fields in the stratified medium were sought, such that the periodically-modulated permittivity and permeability functions could be replaced by constant, averaged, values - the desired effective permittivity and the effective permeability. The effective permeability is disregarded here because the dielectric studied in this thesis (silicon) may be assumed to have the same permeability as free-space at terahertz frequencies [87]. The average field is an accurate representation as long as it only changes slowly in the direction of propagation, which Rytov quantified as

$$k\Lambda|n| \ll 1, \tag{2.18}$$

where $k = \omega/c = 2\pi/\lambda$, Λ is the grating period, and n is the effective refractive index of the medium for the given polarisation and direction of propagation.

The effective properties of the stratified medium may be calculated from the following transcendental equations. The details of the derivation are omitted here, since they may be found in Rytov's paper [86]. The effective permittivity for the TE polarisation, ε_{TE}, is the root of

$$\frac{\alpha_2}{\mu_2}\tan\left(\frac{\alpha_2 b}{2}\right) = -\frac{\alpha_1}{\mu_1}\tan\left(\frac{\alpha_1 a}{2}\right) \tag{2.19}$$

[3]this is the x direction in Rytov's paper [86], where the solution for the TE polarisation (as defined in this thesis) is presented first, under the heading "Propagation along the x axis, e directed along the y axis". The TM solution is presented under the heading "Propagation along the x axis, h directed along the y axis".

where

$$\alpha_1 = k\sqrt{\varepsilon_1 - \varepsilon_{TE}}, \tag{2.20a}$$

$$\alpha_2 = k\sqrt{\varepsilon_2 - \varepsilon_{TE}}. \tag{2.20b}$$

The effective permittivity for the TM polarisation, ε_{TM}, is the root of

$$\frac{\alpha_2}{\varepsilon_2}\tan\left(\frac{\alpha_2 b}{2}\right) = -\frac{\alpha_1}{\varepsilon_1}\tan\left(\frac{\alpha_1 a}{2}\right) \tag{2.21}$$

where

$$\alpha_1 = k\sqrt{\varepsilon_1 - \varepsilon_{TM}} \tag{2.22a}$$

$$\alpha_2 = k\sqrt{\varepsilon_2 - \varepsilon_{TM}}. \tag{2.22b}$$

The transcendental equations do not, by definition, give a closed form solution for the effective permittivities. However, a simple solution may be obtained in the long wavelength limit[4], where the wavelength is much longer than the period, and the arguments of the tangents are small ($|\alpha_1 a|$ and $|\alpha_2 b| << 1$). To a zeroth order approximation, the tangents may be replaced by their arguments, giving the zeroth-order effective permittivities $\varepsilon_{TE}^{(0)}$ and $\varepsilon_{TM}^{(0)}$ as

$$\varepsilon_{TE}^{(0)}(z) = \varepsilon_s f(z) + \varepsilon_i(1 - f(z)) \tag{2.23a}$$

$$\varepsilon_{TM}^{(0)}(z) = \left[\frac{f(z)}{\varepsilon_s} + \frac{1 - f(z)}{\varepsilon_i}\right]^{-1} \tag{2.23b}$$

where TE and TM are the electric field directions specified in Figure 2.3, and $f(z)$ is the fill factor at depth z. Elsewhere, the same result has been derived using classical electrostatic methods, by treating the layers as parallel-plate capacitors that may be added in series (TM) and parallel (TE) [88][5]. Note that this is the difference in the TE and TM effective permittivities for such a 1-D grating that gives rise to the desired form birefringence.

The fill factor $f(z)$ is equal to the duty cycle of the grating at depth z. To illustrate, the fill factors are given for the two example gratings shown in Figure 2.15. The shaded

[4] sometimes called the "quasi-static" limit

[5] this is sometimes referred to in the literature as the "static field" assumption.

areas represent the dielectric material, the unshaded, air. For the lamellar grating in Figure 2.15(a), the fill factor is

$$f(z) = \frac{a}{\Lambda}\Big|_{(0\leq z\leq d)}. \tag{2.24}$$

For the triangular grating in Figure 2.15(b), the fill factor is

$$f(z) = 1 - \frac{z}{d}\Big|_{(0\leq z\leq d)}. \tag{2.25}$$

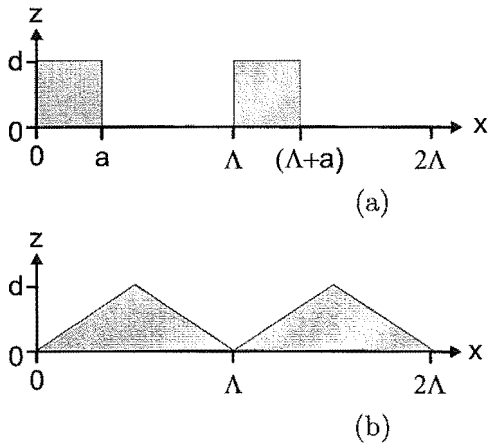


Figure 2.15 Two example grating profiles encountered in the analysis of artificial dielectrics, (a) binary grating, (b) triangular grating. The shaded areas represent the substrate. Two periods (2Λ) of each grating are shown, with (a/Λ) being equal to the duty cycle of the binary grating.

The effective indices of a triangular grating on a substrate, are plotted schematically in Figure 2.16, with the cross section of the structure in Figure 2.16(a), and the calculated indices in Figure 2.16(b). The effective index for the TE polarisation changes linearly with depth z, while the index for the TM polarisation changes non-linearly. If ε_{TE} and ε_{TM} differ in value at any depth, the structure is said to be birefringent. The ability to tailor the birefringence by adjusting the grating profile is useful in the design of artificial dielectric devices such as waveplates, polarisation splitters and polarisation compensators.

Zeroth order EMT is attractive for its simplicity, but its drawback is that it is accurate only in the long wavelength limit $(\Lambda/\lambda << 1)$. Therefore, it must be used with caution for subwavelength gratings with larger periods.

2.4.5 Modelling structures with larger periods

In order to more accurately model subwavelength gratings with a larger period than that suggested by the long wavelength limit, a correction factor may be added to the zeroth order EMT. A power series expansion of the tangents in Equations 2.19

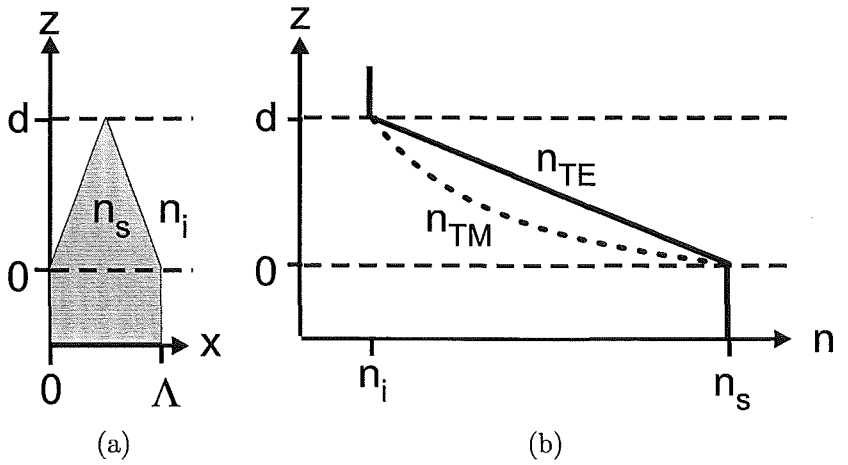


Figure 2.16 A schematic representation of the effective permittivities of a triangular 1-D grating, mounted on a substrate. (a) One period of the triangular grating and substrate. The substrate has permittivity ε_s while the incident medium has permittivity ε_i. The grating period is Λ, and the depth d. (b) The effective permittivities for the TE polarisation (ε_{TE}) and the TM polarisation (ε_{TM}) are plotted schematically for arbitrary values of ε_s and ε_i, for the general case $\varepsilon_s > \varepsilon_i$. The TE effective index varies linearly from the incident value to the substrate value, while the TM effective permittivity varies with the form $1/(a+bz)$, where z is the depth.

and 2.21 produces a second order correction factor in the cubic term of the expansion[6]. The corrected, second order, effective permittivities $\varepsilon_{TE}^{(2)}$ and $\varepsilon_{TM}^{(2)}$ are given by

$$\varepsilon_{TE}^{(2)} = \varepsilon_{TE}^{(0)}\left[1+\frac{1}{3\varepsilon_0\varepsilon_{TE}^{(0)}}\left(\pi(\varepsilon_s-\varepsilon_i)\frac{\Lambda}{\lambda_0}f(z)(1-f(z))\right)^2\right], \tag{2.26a}$$

$$\varepsilon_{TM}^{(2)} = \varepsilon_{TM}^{(0)}\left[1+\frac{\varepsilon_{TE}^{(0)}}{3\varepsilon_0}\left(\pi\frac{\varepsilon_{TM}^{(0)}(\varepsilon_s-\varepsilon_i)}{\varepsilon_i\varepsilon_s}\frac{\Lambda}{\lambda_0}f(z)(1-f(z))\right)^2\right]. \tag{2.26b}$$

Rytov states that Equations 2.26 are valid so long as the correction term is small. Since the equations contain differences of the material parameters, then the region of validity is more limited for structures with vastly different material properties.

More recently, McPhedran and Bell *et al.* investigated the limiting value of Λ/λ for which the homogenisation approach of EMT agreed with rigorous calculations of the exact structure. For both TE and TM the limit was $\Lambda/\lambda < 1/40$ for lamellar gratings, and for arbitrary grating profiles the limit for TE was $\Lambda/\lambda < 1/40$, with a value for the TM limit being unobtainable due to the Fourier expansion of the permittivity profile [70].

[6] $\tan x = x + \frac{x^3}{3} + \frac{2x^5}{15} + \frac{17x^7}{315} + \ldots (|x| < \frac{\pi}{2})$ [89]

2.4.6 Accounting for the depth-wavelength ratio

The solution presented by Rytov assumed that the planar layers were of infinite extent in the in-plane directions. Therefore, the solution is not appropriate for use with shallow gratings. Lalanne *et al.* have shown that the effective permittivities of shallow gratings are strongly dependant on the depth, particularly for depths smaller than about two wavelengths [90]. Additionally, the permittivity of the surrounding layers also has an effect on the effective permittivity for the TM polarisation. Thus, at least for TM, the effective permittivities of shallow gratings must be calculated in the context of the whole problem.

For shallow gratings, Lalanne *et al.* obtained the following expressions for the TE and TM effective permittivities using a Fourier expansion technique:

$$\varepsilon_{TE}(h) = \epsilon_0 + \sum_{p\neq 0} \frac{\epsilon_p \epsilon_{-p}}{2|p|} \left(\frac{\Lambda}{\lambda}\right) hk + O(h^2k^2), \tag{2.27a}$$

$$\varepsilon_{TM}(h) = \epsilon_0 - \sum_{p\neq 0} \frac{|p|\epsilon_p \epsilon_{-p}}{\varepsilon_s + \varepsilon_i} \left(\frac{\lambda}{\Lambda}\right) hk + O(h^2k^2), \tag{2.27b}$$

where h is the grating depth, $\epsilon_p, \epsilon_{-p}$ are the coefficients of the complex Fourier expansion of the permittivity profile, and p is the number of terms included the (complex Fourier) expansion. In the case of gratings with discontinuous profiles, the TM expression becomes infinite because p is in the numerator of the sum. This does not mean that the effective permittivity is undefined for the TM polarisation, but rather that it cannot be obtained using this Fourier expansion method.

In order to illustrate how the effective permittivity changes with depth, Lalanne *et al.* derived an approximate expression for the effective permittivity at arbitrary depths. It was chosen to represent the depth dependance as a function of an arctangent, based on the following three known conditions. Firstly, the permittivity of a zero-depth grating is equal to the zero-frequency component of the Fourier expansion, $\varepsilon(0) = \epsilon_0$. Secondly, the permittivity of an infinitely deep grating is equal to the permittivity given by second order EMT, $\varepsilon(\infty) = \varepsilon^{(2)}$. Thirdly, at intermediate depths, the rate of change of the effective index with respect to depth and wavenumber, is equal to the sum component from the appropriate Equation 2.27, $\mathrm{d}\varepsilon/\mathrm{d}(hk) = \eta_1$, where η_1 is the sum term from Equation 2.27a for TE:

$$\eta_1^{TE} = \sum_{p\neq 0} \frac{\epsilon_p \epsilon_{-p}}{2|p|} \left(\frac{\Lambda}{\lambda}\right) hk \tag{2.28}$$

and the sum term from Equation 2.27b for TM:

$$\eta_1^{TM} = \sum_{p \neq 0} \frac{|p| \epsilon_p \epsilon_{-p}}{\varepsilon_s + \varepsilon_i} \left(\frac{\lambda}{\Lambda} \right) hk. \tag{2.29}$$

Thus, the effective permittivity as a function of depth, $\varepsilon(h)$, may be expressed

$$\varepsilon(h) = \epsilon_0 + \frac{2}{\pi}(\varepsilon^{(2)} - \epsilon_0) \arctan \left(\pi^2 \frac{\eta_1}{\varepsilon^{(2)} - \epsilon_0} \frac{h}{\lambda} \right) \tag{2.30}$$

Lalanne *et al.* validated Equation 2.30 against rigorous calculations, and found good agreement, except for a weak discrepancy in the TM polarisation for depths of about a quarter wavelength.

2.4.7 Effective Medium Theory Selection

Effective medium theory (EMT) may be used to ease the solution of structures containing sub-wavelength gratings (artificial dielectrics), by approximating the inhomogeneous artificial dielectric layer as a homogeneous layer. In the case of 1-dimensionally periodic (1-D) artificial dielectric gratings, the effective permittivity of the homogenised layer is anisotropic, due to the inherent birefringence.

Where the grating is deep, and the period small, and if an approximate solution is sufficient, then Rytov's zeroth-order EMT of section 2.4.4 may be employed in one of two approaches. First, the thin film theory for stacks of homogeneous layers may be used. In the case of lamellar gratings, only a single homogenised layer is necessary. In the case of continuous profile gratings, the grating must first be decomposed into an approximation using tens or hundreds of lamellar gratings, where each lamellar grating is subsequently homogenised. Alternatively, continuous profiles may be treated by the tapered-transmission-line method of Raguin and Morris [83]. If the grating period is larger, it is necessary to employ Rytov's second-order EMT, that provides a correction on the basis of the period-wavelength ratio. The homogenisation approximation has shown to have a critical value of $\Lambda/\lambda = 1/40$, but for gratings with a period smaller than this, EMT may be used in confidence.

Where the grating is shallow ($d < 2\lambda$), the Lalanne solution may be used to account for the strong dependence of the depth on the effective permittivity. The Lalanne solution employs a Fourier expansion of the permittivity profile, that models both lamellar, and continuous-profile, gratings in a single homogeneous layer. This method does account for the period-wavelength ratio. Unfortunately, the use of the Fourier expansion prevents a solution being obtained for the TM polarisation of a grating with a discontinuous profile, and further research is required in this area.

Where an accurate solution is required for artificial dielectrics with large periods and shallow depths, rigorous calculations are preferred. This is worth noting, since

most of the structures studied in this thesis are in this category.

2.5 BIREFRINGENT ARTIFICIAL DIELECTRIC WAVEPLATES

The concept of an artificial dielectric retarder was first introduced to optics by Flanders in 1983 [62]. Flanders was inspired by corrugated-surface twist polarisers (metallic gratings) found on millimetre-wave antennas[7], however it was necessary to adopt a wholly dielectric structure due to the lossy and dispersive nature of metals in the visible.

Flander's 1983 paper reports the fabrication of binary surface-relief gratings having 240nm period. The gratings were constructed from two different materials, polymethylmethacrylate and silicon nitride. The birefringence of the gratings was measured at $\lambda =$ 632.8nm, and was found to agree well, but not exactly, with the birefringence predicted by the simple zeroth-order EMT of Equations 2.23 and the amplitude transmittance approach of Equation 2.9. The good experimental agreement was assisted in part by a modest birefringence and reasonably small index discontinuities at the air-waveplate interfaces (from which reflections were ignored). Furthermore, silicon nitride gratings were used to construct the first known artificial dielectric quarter- and half-wave plates in the visible. In support of the choice of artificial dielectrics in the present work, Flanders states that with the proper choice of material and grating geometry that " ... a wider range [of material properties] may be obtained than is available in the naturally occurring birefringent crystalline materials".

In a similar work, Enger and Case reported the fabrication of waveplates for visible, variously having rectangular and triangular surface-relief gratings, of period 300nm, patterned in photoresist or fused quartz [63] (unlike crystalline quartz, unpatterned fused quartz is not birefringent). The geometry of triangular profile waveplate is shown in Figure 2.17. They too emphasised the importance of the discovery of birefringence in subwavelength gratings, since "This may be especially important in the UV or IR, where suitably transparent and birefringent natural materials may not exist." Due to aspect ratio limitations in the fabrication process, Enger and Crane were unable to etch sufficiently deep to produce a quarter waveplate from a single grating. Instead, they cascaded two to arrive within 1% of a quarter-wave retardance.

Cescato *et al.* investigated quarterwave plates constructed from photoresist gratings having a sinusoidal profile [64]. It was shown that birefringence was still obtainable in gratings with large periods, $\Lambda \geq \lambda$, although the disadvantages were that there was loss due to the presence of diffracting orders, and that a small rotation was required to correct for this loss being unequal between the extraordinary and ordinary rays. Cescato *et al.* used the multilayer approximation of Ono *et al.* [91] to analyse the

[7]Since the twist polarisers are only indirectly related to the present study, they are not reviewed here.

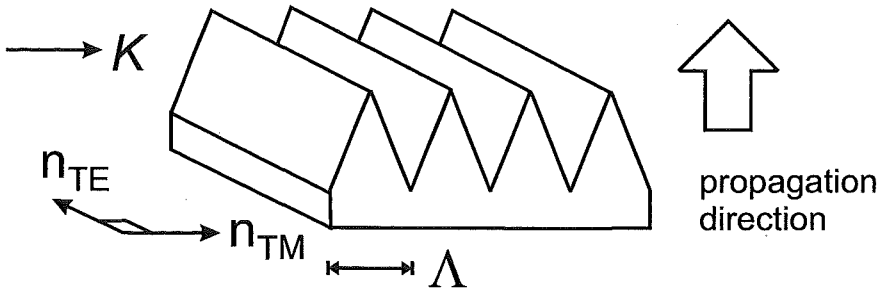


Figure 2.17 An artificial dielectric waveplate having a triangular grating (V-groove) profile. Such plates were reported by Flanders [62] and Enger and Case [63]. For waves travelling in direction of propagation, the effective refractive index seen by wave depends on whether the electric field is aligned parallel to the grating vector $\mathbf{K}$ (n_{TM}), or perpendicular (n_{TE}).

gratings, wherein the continuously graded profile was approximated by many, much thinner, rectangular grating layers each having a different fill factor f.

More recently, Kettunen reported a design for a reflection-mode waveplate in which the reflectivity was boosted to nearly 100% by the use of a multilayer anti-reflection stack underneath a dielectric grating [92]. In terms of reducing loss, reflection-mode devices have the advantage over transmission mode devices, since in practice it is possible to achieve 100% reflection. However, transmission-mode devices, such as the VADR device, have the advantage where a compact system is required, since they may be placed in-line without the need for extra devices to separate counter-propagating beams, or to direct non-orthogonally aligned beams. Since it is envisaged that the development of portable systems may be one of the key drivers of terahertz technology, this is an important advantage. Finally, Kettunen's use of a multilayer stack is also interesting because it shows the kind of optimisation that may be achieved using solely-periodic media.

Importantly, none of these designs provides a variable retardance. At the design stage, the properties of each waveplate may be tailored to suit, which even on its own, is an advantage of artificial dielectrics over naturally birefringent materials. However, the properties are fixed at fabrication and may not be subsequently altered, denying the user the flexibility and convenience associated with variable retarders in the visible, such as the Babinet and Babinet-Soleil compensators. This issue is addressed in the design of a variable artificial dielectric retarder, by Cumming and Blaikie [65], that is further studied in this thesis.

2.5.1 Variable artificial dielectric retarder

The variable artificial dielectric retarder (VADR) device studied in this thesis was proposed by Cumming and Blaikie in 1999, and combines aspects of artificial dielectric waveplates of section 2.5 and the Babinet and Babinet-Soleil compensators of section 2.3. It uses a pair of V-grooved artificial dielectric gratings, similar to those

analysed in the previous section, and is the only variable retarder to be built with artificial dielectrics.

The geometry of the VADR device is shown in the cross section of Figure 2.18, for the case where the two identical interlocking sub-wavelength gratings on the interior surfaces are of the V-groove form. The TE and TM electric field directions are those defined with respect to the grating vector **K** in section 2.2.2, s is the separation between the plates, Λ is the period and d is the groove depth. The operation of the device is described as follows. When the plates are fully interlocked ($s = 0$), there is no birefringence because the device is effectively a uniform slab of dielectric. As the plates are separated an air gap is introduced between the two sets of grooves, allowing the artificial dielectric gratings to emerge. The total birefringence of the device increases with the plate separation distance, s, because the total apparent thickness of the artificial dielectric layers increases, allowing the wave to accumulate a greater phase delay between its components.

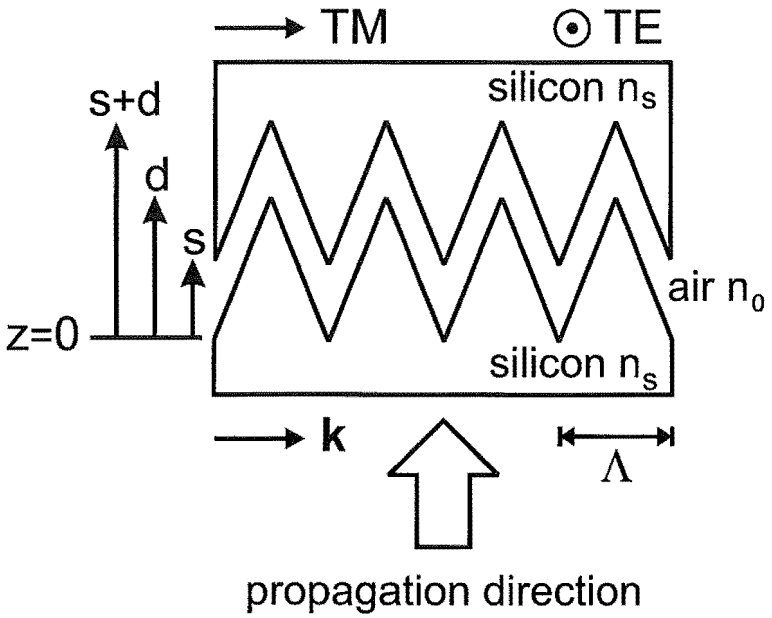


Figure 2.18 The variable artificial dielectric retarder (VADR) device in cross-section. The device is made from two V-grooved substrates with refractive index n_s, groove depth d, groove period Λ (grating vector **K**), and separation s.

The V-groove nature of the grating is not essential; nor is the choice of substrate material (although, as already outlined, silicon is preferable for its high dielectric constant, low loss and low dispersion). Since it is preferable to have subwavelength gratings in order to avoid loss to diffraction, it is unlikely that, with current fabrication and actuation technologies, the VADR could be successfully implemented in the visible or beyond due to the difficulties of precisely fabricating, assembling and actuating a device

with interlocking nanometre scale features. However, it is ideal for terahertz frequencies where there is a lack of naturally birefringent materials having useful optical properties (low loss and low dispersion are desired but unobtainable), and where the required dimensions are well matched to existing microfabrication techniques.

Silicon was chosen as the substrate for the VADR device in the proposal because it has a high refractive index ($n \sim 3.4$), low loss, and excellent dispersion characteristics at terahertz frequencies, as shown in Figure 1.3. Also, the device may be conveniently fabricated by conventional silicon bulk micromachining techniques[8]. For the proposed device, $\Lambda = 50\mu$m, $d = 35\mu m$, and at 1THz the free space wavelength is $\lambda_0 = 300\mu$m and the wavelength in the silicon substrate is $\lambda_s = \lambda_0/n_s = 87.7\mu$m.

As the grating period is sub-wavelength ($\Lambda < \lambda_s$) the retardance Γ of the VADR device can be calculated in the first instance using effective medium theory (EMT). In the proposal, Rytov's second-order EMT of section 2.4.5 was chosen. For the device in Figure 2.18, the depth-dependant fill factor $F(z)$ in the region of the grooves was calculated to be

$$F(z) = \left\{ \begin{array}{ll} 1 - \frac{z}{d} & (0 < z \leq s) \\ 1 - \frac{s}{d} & (s < z \leq d) \\ 1 - \frac{z-s}{d} & (d < z < s + d) \end{array} \right\} \text{ true for} s \leq d \tag{2.31}$$

where z is the depth, s is the plate separation, and d is the groove depth. In the unpatterned substrate $F(z) = 1$ so $n_{TE} = n_{TM} = n_s$, and in the air outside the device $F(z) = 0$ so $n_{TE} = n_{TM} = 1$.

The birefringence of the VADR device can be calculated approximately as follows [65]. The averaged propagation constant that is used to determine the device's phase shift in each region is

$$k_{u,v} = \frac{k_0}{v - u} \int_u^v n(z) dz, \tag{2.32}$$

where v and u define the boundaries of the portion of the artificial dielectric concerned, and k_0 is propagation constant of free space. When $s < d$ the overall phase retardance, Γ, between the TM and TE polarised electric fields can be calculated to be

$$\Gamma_{s<d}(s) = \left(k_{0,s}^{TM} - k_{0,s}^{TE}\right) s + \left(k_{s,d}^{TM} - k_{s,d}^{TE}\right) (d - s) + \left(k_{d,s+d}^{TM} - k_{d,s+d}^{TE}\right) s \tag{2.33}$$

by accumulating the differential phase shifts in the three regions of Equation 2.31.

The proposal [65] goes on to present an improved model that treats the device as a Fabry-Perot cavity with birefringent mirrors. It accounts for the reflections between the plates, and can handle separations exceeding the groove depth. The details are not

[8]The fabrication of the VADR is investigated in chapters three and four of this thesis.

repeated here, however it predicts that once the separation distance exceeds the groove depth, the retardance oscillates as a function of separation distance.

Since the ratio Λ/λ is greater than the critical value of 1/40 for the proposed VADR, EMT could not be used with complete confidence. Therefore, a numerical vector Maxwell Equation solving tool was also used. The particular tool employed was the multiple-multipole method (MMP), which is described further in the comparison of simulation techniques presented in Chapter 6.

The calculated retardance is plotted in Figure 2.19, along with the retardance predicted by the zeroth order EMT and Fabry-Perot cavity model (labelled $EMT^{(0)}+FP$). In the fully intermeshed position ($s = 0$) there is no retardance ($\Gamma = 0$). As the separation is increased, the retardance rises rapidly. The plates make the transition from being partially interlocked to being fully separated at $s = 35\mu$m, as indicated, but the retardance continues to rise until it reaches a maximum at $s = 70\mu$m. Thereafter, the retardance oscillates in the manner of a Fabry-Perot cavity, with the period of oscillation (in terms of separation) being $s = 150\mu$m, corresponding to a half wavelength.

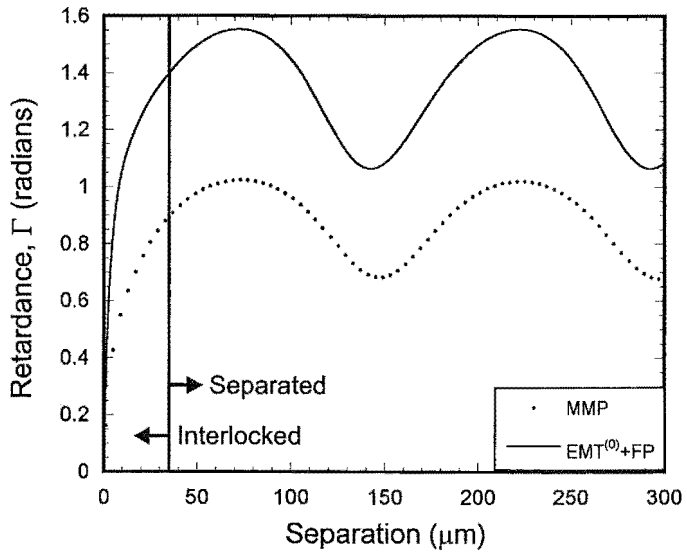


Figure 2.19 Predicted retardance of the variable artificial dielectric retarder (VADR) device for design frequency of 1 THz. $EMT^{(0)}$-FP: Semi-analytical zeroth-order Effective Medium Theory (EMT) - Fabry-Perot cavity model, MMP: Multiple-Multipole numerical vector Maxwell Equation solving tool. As indicated, the plates are interlocked for $0 \leq s < 35\mu$m, and separated for $s > 35\mu$m.

It may be seen that the zero-order EMT-based solution predicts a much greater maximum retardance than MMP (at $s = 0, \Gamma_{EMT^{(0)}+FP}$ =1.55 radians compared to $\Gamma_{MMP} = 1$ radians). The apparent differences between the MMP and the EMT results arise from inaccuracies in the EMT approximation. It is identified in the proposal that the large period and high dielectric constant are contributing factors. It may also be seen from section 2.4.6 that the shallow grating depth is another.

It should also be noted that the analysis in the proposal assumed that the unpatterned rear surfaces of the VADR device would be augmented with an anti-reflection artificial dielectric surface or dielectric quarter-wave coating. The latter is often used with uniaxial crystal waveplates in the visible [77], because it reduces the mismatch in the TE and TM transmission magnitudes. The effect of this assumption on the VADR device analysis in proposal was that the reflections from these outer surfaces could be neglected. This assumption is often made in the analysis of other artificial dielectric retarders.

A large portion of this thesis is devoted to a practical realisation of a VADR device. This includes the development of a microfabrication process, measurement and analysis of the performance, and design of improved devices.

2.5.2 Other birefringent retarders

Artificial dielectrics, and 1-D gratings in particular, are not the only structures capable of producing a useful birefringence. For example, the waveplate investigated by Lie-Ming Li [93] comprised a 2-D array of parallel dielectric cylinders, the axis of the cylinder being orthogonal to the directions of periodicity, as shown in Figure 2.20. The arrangement of the cylinders into a square lattice is indicated by the unit cell (in the upper-left corner of the figure) while the arrow indicates the wave is incident along a direction orthogonal to the axis of the cylinders. This structure is not an artificial dielectric because the period is sufficiently large that diffraction occurs, and is in fact a type of photonic band gap structure.

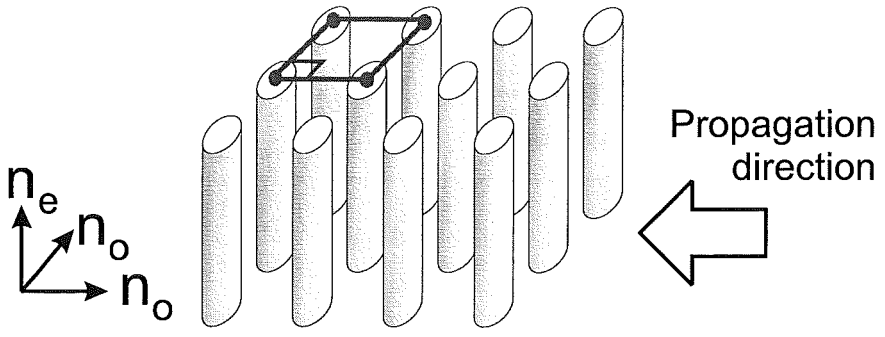


Figure 2.20 Lie-Ming Li's two-dimensionally periodic (2-D) photonic crystal waveplate concept [93]. A 2-D periodic array of dielectric cylinders exhibits birefringence.

Two main advantages may be obtained by using photonic band gap structures (or artificial dielectrics, for that matter) as retarders. Firstly, a large birefringence may be generated, resulting in a more compact device. In this instance, to obtain a half-wave retardance, Li's waveplate need only be 0.7λ thick, compared to the 2.9λ thickness required for a calcite waveplate. Secondly, any dielectric with a low loss and low dispersion may be used. This is important because there is yet to be found a naturally

birefringent material at terahertz frequencies with these properties. A third possible advantage proposed by Li, relating only to photonic crystals and not to artificial dielectric retarders with only one or two gratings, is that photonic band gap materials could be tuned to provide equal transmission coefficients for both polarisations, although at this time, his work on this particular matter has not been published.

Photonic crystals are of interest in this thesis for an altogether different application, a tunable frequency selective filter, where birefringence is not required. This chapter now goes on to provide background to the tunable frequency-selective filter, and does not further deal with birefringence or polarisation devices, except for in the summary at the end of the chapter.

2.6 PHOTONIC CRYSTALS FOR TUNABLE FILTERS

2.6.1 Introduction

Photonic crystals are ideal for the construction of bandstop and bandpass frequency-selective filters in both transmission and reflection modes of operation, due to the characteristic gaps in their electromagnetic dispersion equation. Practical advantages that may be realised by photonic crystal filters include wide bandwidth, strong attenuation in the stop band, strong cutoff at the edges of the band, and an ability to have, optionally, one or more narrow pass-bands located within a much wider stop band.

Furthermore, the filter properties of photonic crystals may be readily tuned during design (before fabrication). However, it is much more difficult to tune the filter after it has been fabricated, so that it may be used "on-the-fly" to, for example, shift between communications channels in a transceiver, or shift between chemical absorption bands in a spectrometer. In this study, it is desired to create just such an "on-the-fly" tunable filter using photonic crystals. This topic is also of interest to other researchers, but many studies are still in early stages, and reference here will only be made to a few that are most closely related to the present work.

2.6.2 History

Photonic band gap (PBG) structures (photonic crystals) were first proposed in 1987 for the purpose of inhibiting spontaneous photon emission in semiconductors [94, 95]. It was desired to find a three-dimensionally periodic structure that would prevent the propagation of light in any direction, for a given band of frequencies (band gap). This goal of obtaining a full 3-D band gap was reached some four years later [96], in good agreement with concurrently-developed theory [97]. The structure was known as Yablonovite (named after its creator), and it is shown in Figure 2.21. Since then, researchers in the field have continued to study crystals with 3-D gaps, but a gradually increasing amount of attention has been paid to other applications such as filtering,

wave guiding [98] and dispersion compensation [99], where a full 3-D band gap is not required. In these applications, photonic crystals offer advantages such as strong stop band attenuation, lossless bends (in waveguides), compactness, ease of fabrication and the ability to tailor the optical properties.

2.6.3 Theory

Photonic crystals exhibit a characteristic (band) gap in their electromagnetic dispersion relation. Prior to work arising from, and inspired by, the 1987 proposals of Yablonovitch and John [94,95], such behaviour was only known for 1-D structures such as multilayer reflectors, as used in optics. Creating a band gap for all directions of propagation implied that 3-D modulation of the refractive index was required. Considerable effort was required to develop a method of analysis that would readily and confidently reveal whether the photonic band gap was present for *all* directions of propagation, but without needing to solve for every single possible direction of propagation (potentially an infinite set).

Initial inspiration came from solid state physics, where theoreticians were already familiar with solving for electronic band gaps in the 3-D lattices of semiconductors. Electronic band gaps were identified by considering periodic variations in potential (a scalar field) within the microscopic semiconductor lattice. With this in mind, the first attempts to solve for photonic band gaps used plane wave expansions of the electromagnetic mode structures in the reciprocal crystal, and made the assumption of scalar electromagnetic waves [100–102]. Yablonovitch predicted the assumption of scalar electromagnetic waves would give invalid solutions for many of the cases of interest, because it was not possible to accommodate electromagnetic mode structures having different polarisations [103].

Simultaneously, Yablonovitch reported that he had experimentally obtained a full band gap for a structure resembling a stack of chequerboards, and having face-centred-cubic symmetry [103]. Shortly thereafter, the theoreticians had improved their analysis to deal with vector electromagnetic waves and were able to analyse Yablonovitch's structure with increased confidence. Unfortunately, upon analysis with the new full-vector method, it was discovered that a full photonic band gap could not be attributed to Yablonovitch's chequerboard structure, there being an observed band gap at one point (the W-point) where none was predicted in theory [97,101,104]. These works prompted some speculation that the concept of full photonic band gap was not achievable [105]. Fortunately, this was not the case, since further studies with the full-vector method predicted that a diamond structure would have a full band gap [97,106]. Yablonovitch confirmed this by experiment, when he demonstrated a full band gap in the eponymous Yablonovite shown in Figure 2.21 [96]. This marked the successful development of an analysis method, derived from Maxwell's equations, that could accurately predict full

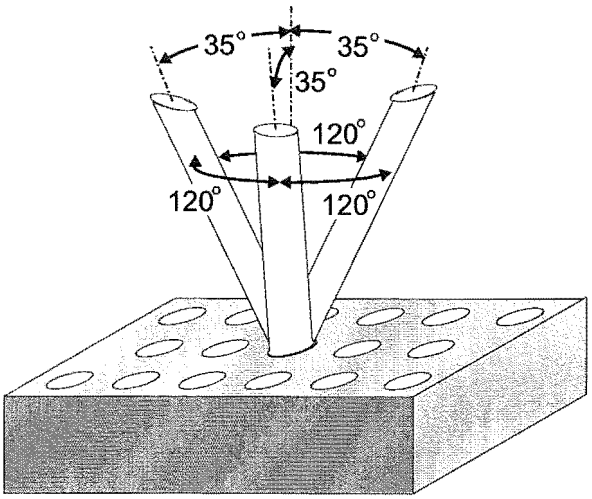


Figure 2.21 Yablonovite was the first demonstrated photonic crystal to have a full three-dimensional band gap. A diamond-like lattice structure was achieved in the dielectric, by drilling each hole three times at an angle of 35.26° away from the normal, and spread 120° on the azimuth. The holes are arranged in a triangular lattice. For microwave wavelengths, a real drill bit can be used, or for higher frequencies, reactive ion etching techniques may be used.

3-D photonic band gaps.

Before briefly describing the full-vector plane-wave-expansion method used to predict the 3-D photonic band gap in diamond structures, it is worth emphasising again that photonic band gaps are an electromagnetic effect completely predicted by Maxwell's equations. Therefore, many solutions to Maxwell's equations, and not just the full-vector plane-wave-expansion mentioned above, are potentially capable of predicting photonic band gaps. It was simply that those searching for a 3-D band gap desired a method that allowed them to readily predict with confidence whether the photonic band gap existed for *all* directions of propagation. Therefore, in circumstances where it is not desired to obtain a full 3-D band gap, other solutions to Maxwell's equations may be more appropriate.

This may be illustrated by taking a 1-D example. A multilayer stack is often used in optics because it is capable of 100% reflectivity at the design wavelength. It may be shown that, at least for an arbitrary example, the high-reflectivity at and near the design wavelength corresponds to a photonic bandgap. One such example is presented in Figure 2.22, with a schematic of the multilayer stack in Figure 2.22(a), an analysis of the reflectivity in Figure 2.22(b) and an analysis of the photonic band structure in Figure 2.22(c). The multilayer stack considered here comprises alternating layers of air ($n_1 = 1$) and GaAs ($n_2 = 3.6$), each having equal thickness $a/2$, and infinite lateral extent. In the optical calculation, a finite number of layers were required for the calculation. Somewhat arbitrarily, it was chosen to have 14 GaAs layers, 13 air layers, plus two additional semi-infinite air layers at each end of the stack. In the photonic band structure calculation, it was assumed that there was an infinite number of layers.

The incident wave is directed along the normal to the stack's front surface. In the following, it is first solved for the reflectivity, then for the photonic band structure. The features of the photonic band structure diagram may not be immediately obvious, so these will be explained in due course. For now, it is sufficient to say that both methods predict no transmission through the multilayer stack for the frequencies within the shaded band.

Physically, if a monochromatic plane wave is allowed to propagate through the stack, there will be multiple reflections from each of the dielectric interfaces between the layers. The many reflected waves will interfere, and if they do so destructively, then transmission through the stack will be prohibited. Thus, if the materials are lossless, none of the energy is absorbed in the stack and there will be 100% reflection. In optics it is unusual to track each of the reflected rays explicitly, since there are infinitely many. Instead, total electromagnetic field quantities may be considered, and thus all reflections may be accounted for. There are several mathematical treatments, and for Figure 2.22(b) the T-matrix method that is presented in section 6.2.1 (see also Hecht and Zajac [80]) has been used. Since the derivations are not sufficiently compact, they are not repeated here.

The full-vector method used to produce the photonic band diagram of Figure 2.22(c) is now briefly described. With this method, the electromagnetic mode structures of a macroscopically structured, linear, lossless dielectric material (such as the 1-D stack of Figure 2.22a) may be determined "ab-initio" from Maxwell's equations. An introductory text presents, with commentary, the derivation of the master equation [107] which must be solved in order to determine the band structure, while details of various computational schemes for obtaining the solution may be found in literature, for example [108][9]. Ordinarily, the master equation is developed in terms of the magnetic field for the convenience of the iterative eigensolving routines in the computational code; the master equation may be derived according to procedure in Appendix A.3, or [107], to be

$$\nabla \times \left(\frac{1}{\varepsilon(\mathbf{r})} \nabla \times \mathbf{H}(\mathbf{r}) \right) = \left(\frac{\omega}{c} \right)^2 \mathbf{H}(\mathbf{r}), \tag{2.34}$$

where $\varepsilon(\mathbf{r})$ is the dielectric constant within the photonic crystal as a function of the position $\mathbf{r}$. This master equation may be treated as an eigenvalue problem, with field patterns of the modes as the eigenvectors $\mathbf{H}(\mathbf{r})$ and the frequency of the mode being proportional to square root of the eigenvalues $(\frac{w}{c})^2$.

It should be noted that the exact reflectivity of the stack is not predicted by the band structure plot – it simply indicates the region for which transmission is prohibited (maximum reflectivity). In many engineering problems, the reflectivity and transmissivity are important. Therefore, in these cases, it is useful to use alternative

[9]This code freely available, see reference for details.

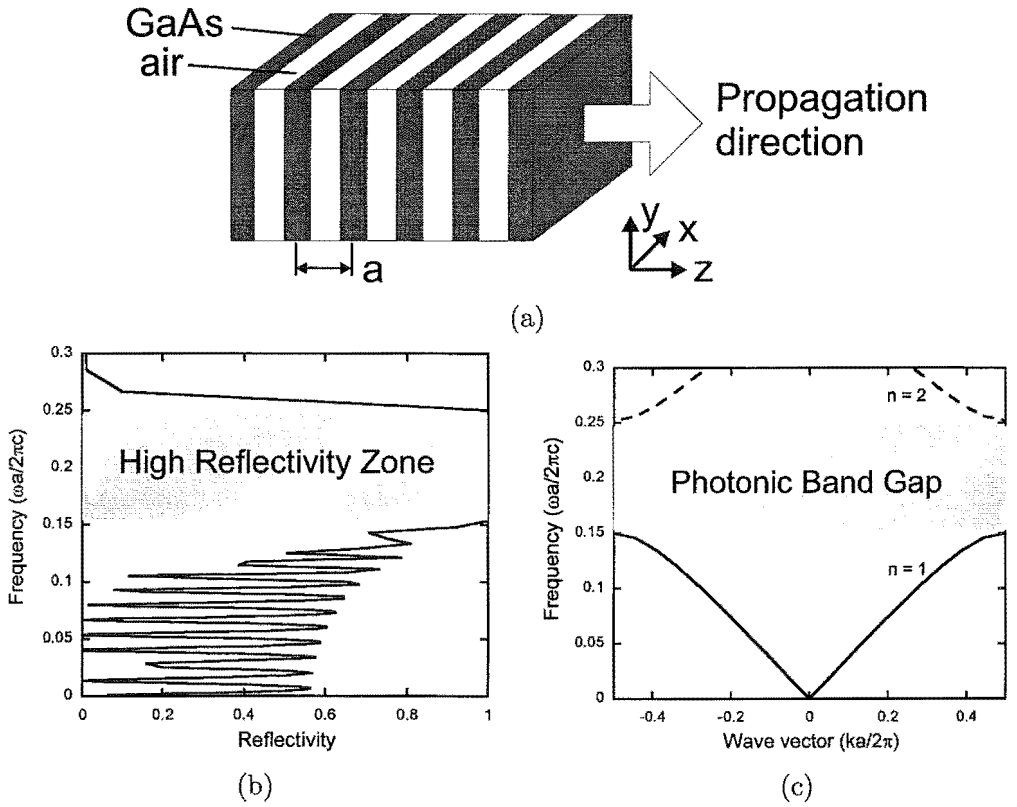


Figure 2.22 A multilayer stack of GaAs ($\varepsilon = 13$) and air ($\varepsilon = 1$), with each layer having thickness $0.5a$ is analysed. A 27-layer stack was considered. (a) stack geometry (showing only 11 layers), indicating direction of propagation and the stacking period a. Analysing the reflectivity using typical optical methods gives the plot in (b), while analysing the photonic band structure directly gives the plot in (c). A photonic band gap is present between frequencies of dimensionless value 0.15 and 0.25, indicated by the shaded bands in (b) and (c).

analysis methods that provide the reflectivity and transmissivity. If this information is available, it is not necessary to independently and explicitly solve the photonic band structure, since photonic band gaps can be identified by examining the reflectivity or transmissivity of a structure. For example, examining Figure 2.22(b,c) it may be seen that the high reflectivity region and the photonic band gap correspond to identical frequency ranges and that therefore the reflectivity (or transmissivity) plots may be used to identify photonic band gaps. The ability to determine photonic band gaps using other structures is also important when the structure to be analysed has a discontinuous permittivity profile $\varepsilon(\mathbf{r})$, because this can give rise to slow convergence and limited accuracy when using the plane wave expansion method [106].

The only other method to explicitly calculate the band structure is the transfer matrix method (TMM)[10] developed at Imperial College. The TMM has the advantage of being able to calculate the reflection coefficient as well as the band structure (whereas the plane wave expansion method can only calculate the band structure). It has also been shown to be slightly more efficient, due to the use of a different expansion of the mode structure. The method is well described in Pendry's paper [109], but it is not described in any more detail here since it is not used in the thesis. At present, both the original code and a more recent derivative (the latter produced at the University of Glasgow) are freely available [110, 111].

Other numerical techniques that have been applied to the study of photonic crystals include the Finite Element Method (FEM) [112, 113] and the Finite Difference Time Domain (FDTD) method. Both techniques are described in more detail in section 6.2.

The final point to be mentioned in this section is that the scaling property of Maxwell's equations has useful implications for practical research into photonic crystals. Since there is no fundamental length scale associated with the electromagnetic properties of a macroscopic system, it is possible to scale a device by expanding and contracting all the dimensions, including the wavelength. Thus, the solution at one scale length can be used to determine the solutions at all scale lengths. Therefore, devices may be prototyped at a convenient, perhaps lower, frequency. In this way, microwave experiments may be used to determine the properties of a particular crystal structure for all wavelengths.

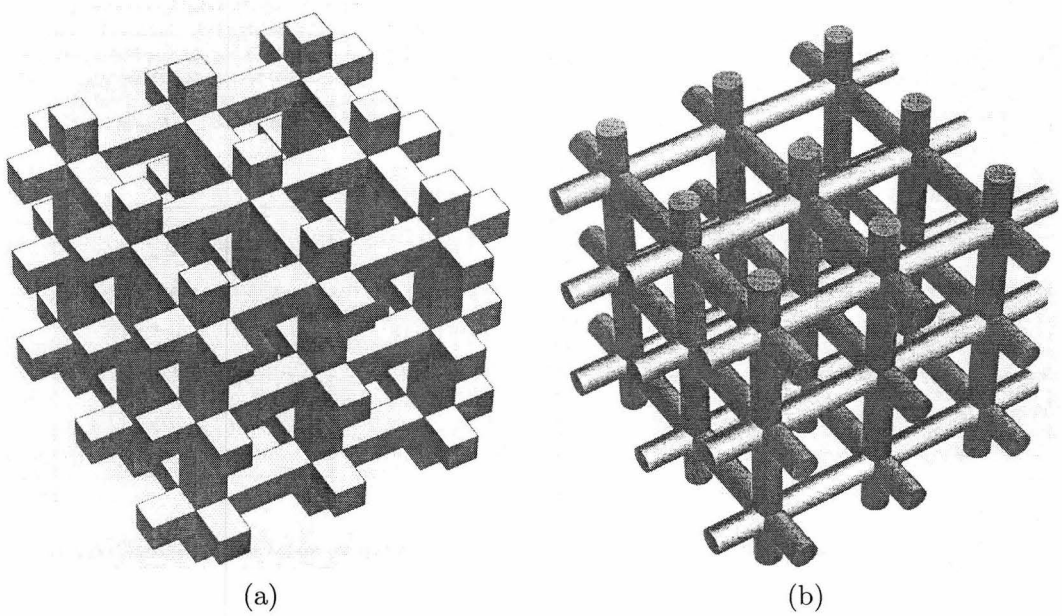
2.6.4 The scaffold and wood-pile crystals

The tunable filter investigated in this study has a simple tetragonal (ST) lattice structure, however it is similar in appearance to both the face-centered-tetragonal (FCT) woodpile photonic crystal [114] (see Figure 1.1(b) for the FCT woodpile's unit cell), and in a special case (when length, width and heigh of the simple tetragonal unit cell are equal) it shares the simple-cubic (SC) lattice structure of the scaffold crystal [115].

[10] This method is not to be confused with the similarly named T-matrix method

Therefore, the properties of both of these crystals are described. The scaffold structure was proposed in 1993, one year before the woodpile, so it is with the scaffold structure that this description begins.

The scaffold structures are shown in Figure 2.23, with the square rod scaffold in Figure 2.23(a) and the circular rod scaffold in Figure 2.23(b). They are based on the simple cubic lattice of Figure 2.24, with the simple cubic lattice in Figure 2.24(a) and the reciprocal lattice (Brillouin Zone (BZ)) in Figure 2.24(b). The simple cubic lattice has four symmetry points, one in each corner. It has been shown in the computation of other band structures that the topology of the structure is important in determining the properties of the crystal, by which it is meant that structures having the same lattice symmetry will not necessarily have the same properties if the crystal has a different form within that lattice.



(a) (b)

Figure 2.23 The scaffolded structure may be constructed with square rods (a), or circular rods (b).

The advantage of the scaffold structure is that it is simple, and therefore easier and more economical to fabricate, as compared to say, the more complicated Yablonovite of Figure 2.21. Interestingly, in the square-rod scaffold of Figure 2.23(a), the air and the dielectric regions have exactly the same geometry at all non-trivial fill factors ($f \neq 0, 1$), save for the relative size being different for fill factors $f \neq 0.5$.

In the square-rod scaffold, the largest complete gap was found for a fill factor$f$ = 0.82, where the dielectric occupied 82% of the volume of the unit cell. The photonic band structure is plotted in Figure 2.25 for case where the dielectric is GaAs (n_{GaAs} = 3.6) and the other medium is air ($n_{air} = 1$). A full band gap is present between the second and third bands for wavelengths approximately in the range $0.66\frac{2\pi}{a} < \frac{\omega}{c} <$

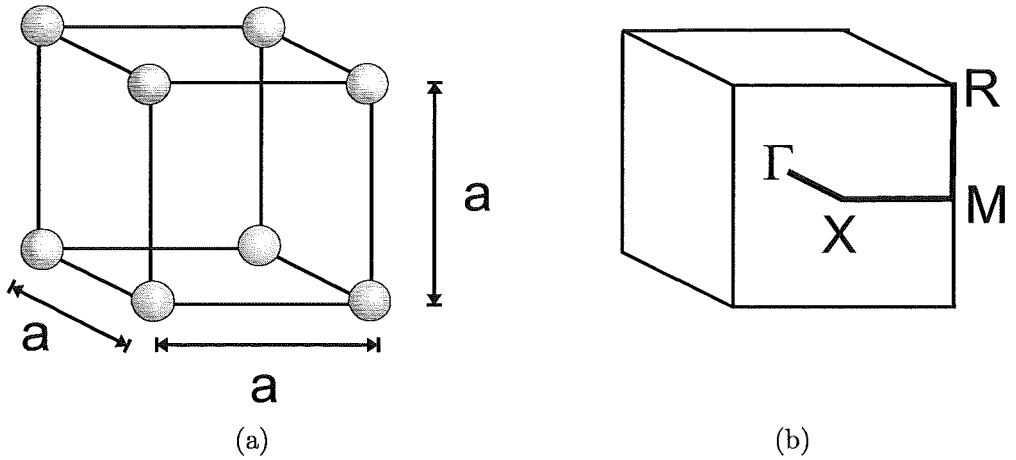


Figure 2.24 The scaffolded structure has the symmetry of a simple cubic (sc) lattice. (a) The sc lattice measures one lattice constant a in each of the three orthogonal directions. (b) The Brillouin Zone (BZ) of the sc lattice is also a cube, and the four main directions in the irreducible BZ are labelled. Γ is located at the centre of the cube, X in the centre of the face, M in the centre of the edge, and R in the corner

$0.71\frac{2\pi}{a}$, where $\bar{c} \equiv c/\sqrt{\bar{\varepsilon}}$, where $\bar{\varepsilon}$ is the spatial average of $\varepsilon(\mathbf{r})$.

The woodpile structure is shown in Figure 2.26 for the case of square rods, although it may be constructed with elliptical or circular rods. The woodpile comprises layers of parallel rods in which the stacking sequence repeats every four layers. Within each layer, the rods are parallel, and spaced by a. Between layers, the rods are rotated by 90°, while the second set of two layers is shifted by $0.5a$. It was proposed by Ho *et al.* in 1994, and like the scaffold, it is simple in construction. Both structures are well suited to microfabrication at terahertz frequencies.

The woodpile has a face-centred-tetragonal (FCT) lattice symmetry as shown in Figure 2.27. This is similar to the simple cubic lattice, except there is an extra "atom" at the centre of the unit cell, and the height is variable, now equal to c. The BZ for this lattice structure is somewhat more complicated, and is not drawn here. Instead, it is shown in Figure 2.27 how the fct lattice of Figure 2.27(a) degenerates into a face-centred-cubic lattice for the special case of $c/a = \sqrt{2}$, as shown in Figure 2.27(b).

The band structure of the woodpile is plotted in Figure 2.28. There are more wavevector directions indicated because the woodpile crystal's FCT BZ has more lines of symmetry than the SC BZ of the scaffold. In this case, the structure is constructed from rods of GaAs ($n_{GaAs} = 3.6$), while the other medium is air ($n_{air} = 1$). The fill factor is 26.6% (considerably less material is required than for the scaffold), and the ratio of the fct lattice is $c/a = 1.22$. The midgap frequency is inversely proportional to lateral dimension a, therefore larger structures give a lower frequency bandgap. But in order for a gap to open up in the first place, Ho *et al.* showed that a minimum refractive index contrast of $n_2/n_1 > 1.9$ was required. This is in contrast to a simple

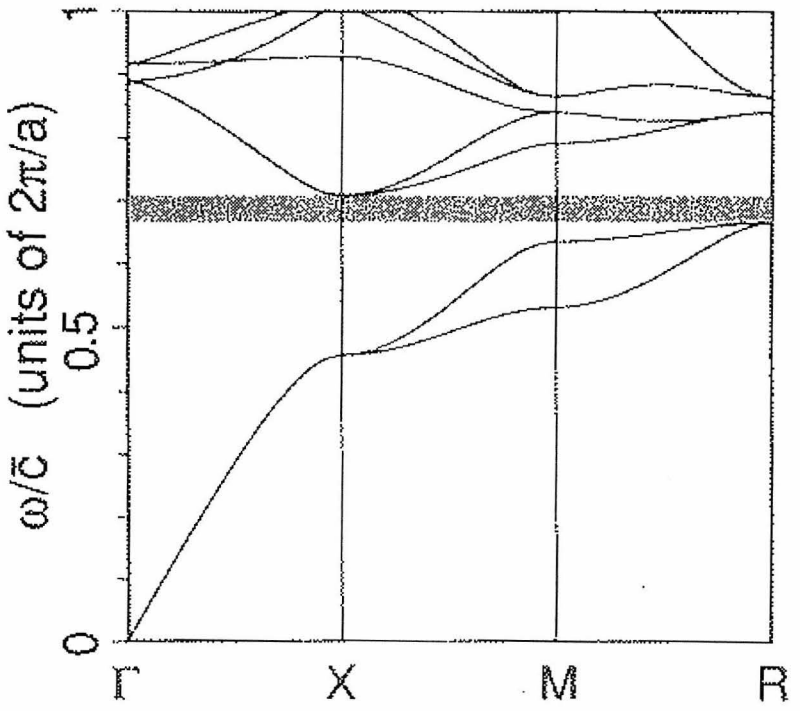


Figure 2.25 The band structure for the simple cubic scaffold structure having square rods of GaAs ($n_{GaAs} = 3.6$) interspersed with air $n_{air} = 1$, with a fill factor of $f = 0.82$. The first band gap lies between the 2nd and 3rd bands (shaded). Note that $\bar{c} \equiv c/\sqrt{\bar{\varepsilon}}$, where $\bar{\varepsilon}$ is the spatial average of $\varepsilon(\mathbf{r})$ [115].

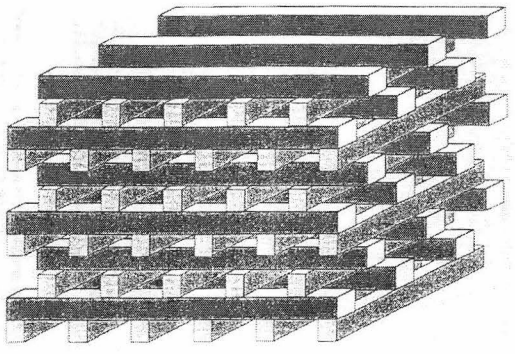


Figure 2.26 The woodpile structure may be constructed with square rods as shown, or with elliptical or circular rods [114].

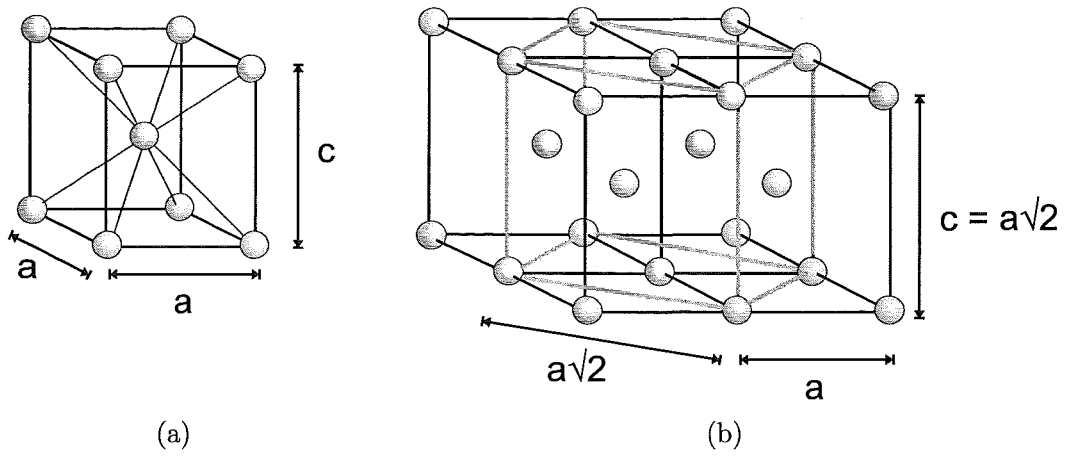


Figure 2.27 The woodpile structure has the symmetry of a face centred tetragonal (FCT) lattice, except for one special case. (a) The fct lattice measures one lattice constant a in the two lateral directions, while in the vertical direction the lattice constant, c, is different. (b) For the special case where $c = a\sqrt{2}$, the fct lattice degenerates into a face-centred-cubic lattice having the orientation and dimensions indicated.

multilayer stack, for which a narrow band gap opens up almost immediately. Therefore, for woodpile crystals in air, the dielectric substrates must have a refractive index greater than 1.9. Silicon ($n_{Si} \sim 3.4$) would be a suitable choice of dielectric substrate for this type of structure at terahertz frequencies.

Several other properties provide useful benchmarks for tunable photonic crystal filters developed in this study, particularly with regards to stop band width, and rejection. The maximum gap to mid-gap ratio, a measure of the bandwidth of the stop band, is about 18% for the square-rod woodpile but falls dramatically for fct lattice ratios $c/a < 1.5$ or $c/a < 1.5$, resulting in a much narrower bandgap. The gap-midgap ratio is also affected by the filling factor, although the optimal filling factor ($f \sim 0.30$) depends in turn upon the refractive index contrast and topology. In each case of interest, the dependance may be found empirically, by simulating many structures having the same refractive index contrast and topology but each having a slightly different fill factor.

The attenuation of a 4-layer unit cell was found to be, at mid gap frequencies, 21dB (theory) and 17dB (experiment, microwave regime). This is important because it indicates that not many layers are required for effective rejection of unwanted electromagnetic signals. For example, 8 layers (2 unit cells) would give 42dB which Ho *et al.* claim is large enough for many applications [114].

Since the woodpile was first investigated, it has been demonstrated over most of the lower half of the terahertz frequency range, as indicated in Table 2.3. It may be seen that the band gaps achieved are reasonably broad, which is useful for rejecting a broad band signal. However, filters with much narrower bands also have their uses, for

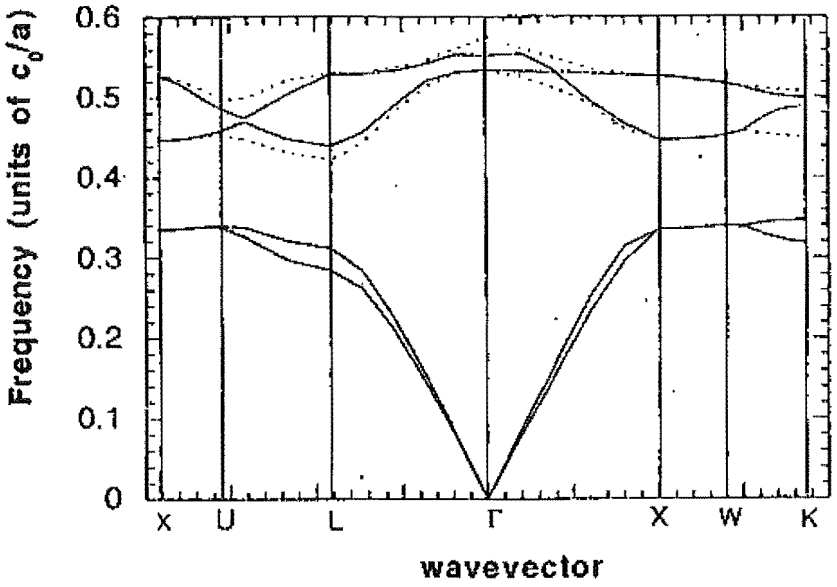


Figure 2.28 The band structure for the face centred tetragonal (FCT) woodpile structure, having square rods of GaAs ($n_{GaAs} = 3.6$) interspersed with air $n_{air} = 1$, with a fill factor of $f = 0.266$. The first band gap lies between the 2nd and 3rd bands, but is not shaded. The ratio of the fct lattice is $c/a = 1.22$. Note that c_0 is the velocity of light in a vacuum, and that the dimensionless frequency on the vertical axis is referenced to the lateral dimensions of the fct lattice, a [114].

example selecting between adjacent closely-spaced channels. The next section describes how a narrow passband may be created within the stop band of a photonic crystal.

2.6.5 Crystals with defects

The introduction of one or more carefully controlled imperfections (defects) into the crystal structure results in the creation of a narrow pass band within the stop band [120, 121]. This is of use in the creation of filters, since narrow band filters are as useful, if not more, than a broad band filter. The defect-mode filtering concept is illustrated in Figure 2.29 for an arbitrary and somewhat simplistic example, where the incident signal in Figure 2.29(a) has two sidebands that may be filtered out by passing the signal through a filter having approximately the properties illustrated in Figure 2.29(b),

Table 2.3 Woodpile crystal experiments. The stop band attentuations are reported in units of dB/cell, where a cell is four layers of rods. (n.r. = measurement not reported.)(* theoretical)

Experiment frequency	Band gap frequency	Stop band attenuation	Researchers	Year
n.r. (microwave)	n.r. – n.r.	16dB/cell	Ho *et al.* [114]	1994
70 – 120 GHz	81 GHz – n.r.	16dB/cell	Ozbay *et al.* [116]	1994
250 – 550 GHz	425 – 518 GHZ	17dB/cell*	Ozbay *et al.* [117]	1994
100 – 600GHz	205 – 315 GHz	n.r.	Chelnokov *et al.* [118]	1997
1.4 – 5.0 THz	2.0 – 3.0 THz	12dB/cell	Lin *et al.* [119]	1998

thus giving only the desired component of the signal at the output, as illustrated in Figure 2.29(c).

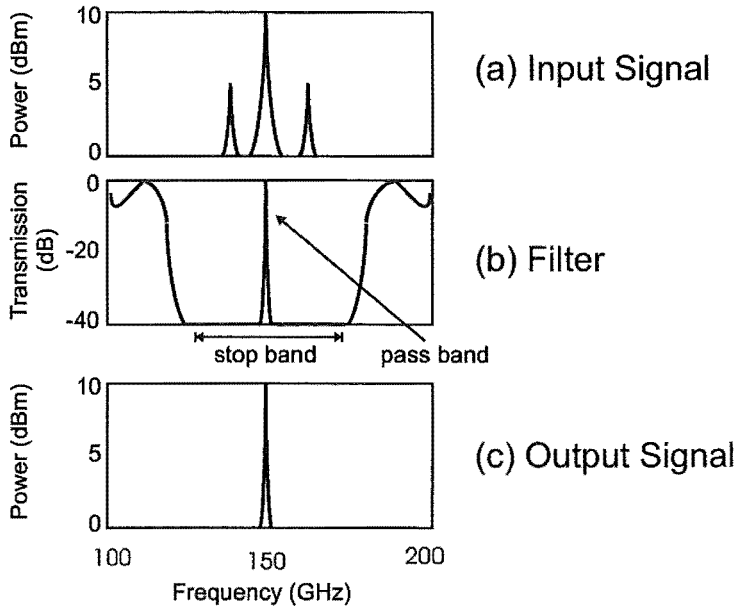


Figure 2.29 A useful filter could be created by introducing a narrow transmission peak into the broad stop band of a photonic crystal. For example, (a) an incident signal with one set of unwanted sidebands could be introduced to (b) a filter having a narrow pass band within a broad stop band, giving at the output (c) the desired signal.

Returning to the 1-D example, a defect may be introduced to a multilayer stack by enlarging or reducing one of the layers, such as shown in Figure 2.30. The middle air layer has been enlarged, giving the familiar optical structure of a Fabry-Perot etalon, surrounded by two multilayer mirrors.

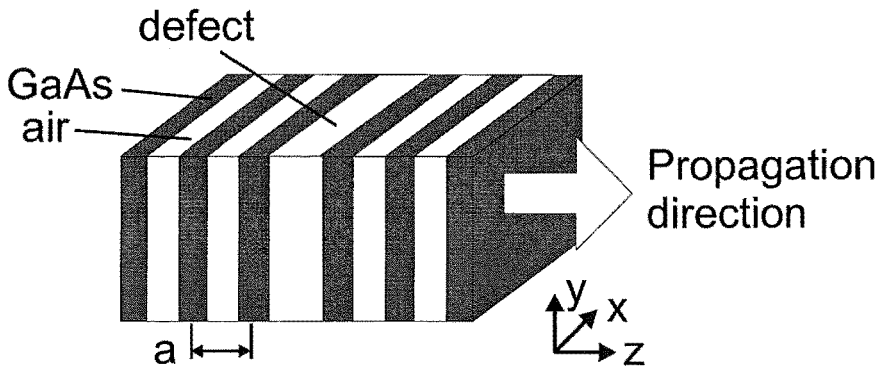


Figure 2.30 The GaAs multilayer stack with an enlarged air layer in the middle of the stack. In optical terms, this is a Fabry-Perot etalon, but in terms of photonic band gaps it is a disturbance to the periodicity, and hence a "defect".

The presence of a defect in the stack allows previously prohibited modes to exist. This may be illustrated by considering the density of states, which, for photonic band

gaps, is equivalent to the number of different modes (field distributions) that may be supported at a given frequency. The estimated density of states diagrams are shown in Figure 2.31, for the structure without the defect in Figure 2.31(a), and for the structure with the defect (enlarged air layer) in Figure 2.31(b) [107]. A wave may only propagate through the crystal if there is a non-zero number of modes supported at its frequency, otherwise, it will decay evanescently. Within the photonic band gap of the structure with no defect, zero modes are supported and therefore transmission is prohibited. In the photonic band gap of the structure with the defect, many previously-prohibited modes are supported by the defect at a particular frequency, and transmission at the frequency of the defect modes may be possible.

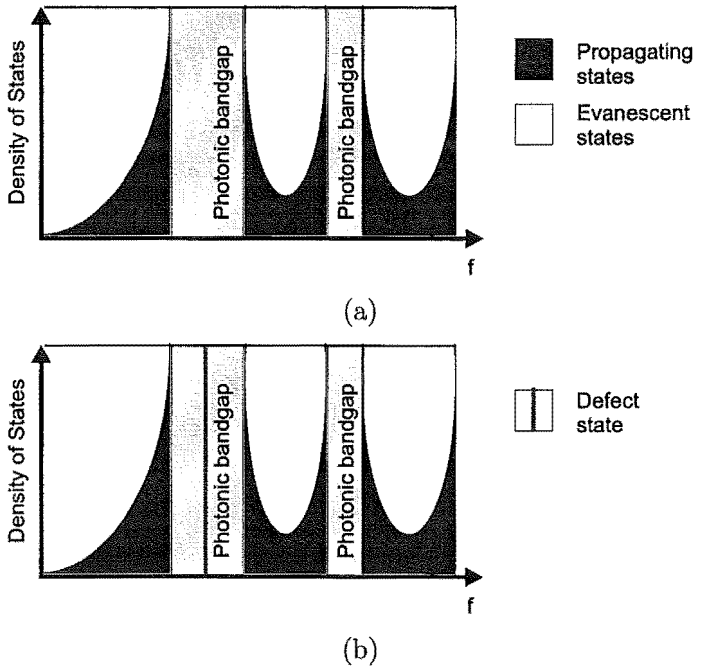


Figure 2.31 Estimated density of states diagram for a multilayer stack having (a) no defect (b) defect. When the multilayer stack has a defect, there is a high density of states for frequencies associated with the defect. In photonic band gaps, the density of states corresponds to the number of possible modes (field structures) that may be formed. Diagram after [107].

A wider range of defects may be created in 2-D and 3-D crystals, as compared to 1-D crystals. In the case of 1-D structure, a defect can only create a localised mode in a plane, whereas in a 2-D structure, the mode may be localised in lines. In a 3-D structure, the mode may be localised in a point, analogous to a resonant microcavity. Defects in 2-D are interesting because they may be used to create waveguides [98], while point defects in 3-D are able to enhance spontaneous emission, because of the large density of states associated with the defect [122].

In the context of a filter, the problem with using a defect to create a narrow passband is that propagation of the defect modes is not supported elsewhere in the

crystal. Thus, a wave at the defect frequency, incident on the crystal from outside, will decay evanescently until it reaches the general region around the defect, where it may propagate. Upon reaching the other side of the defect region, it must then penetrate through to the outside of the crystal in order to be observed. Even if it reaches the outside of the crystal, the signal must have been necessarily attenuated as a result of evanescent decay in the undisturbed parts of the crystal. Therefore, if it is desired to create a filter with good transmission in the narrow passband associated with the defect, it is important to have good coupling between the defect mode and the outside of the crystal.

However, the practical difficulty of achieving good coupling of the defect mode in a dielectric crystal is evident in the various experimental studies. For example, Figure 2.32 shows the transmission through a crystal structure, having no defect in Figure 2.32(a), a single, centred, donor defect in Figure 2.32(b) and a single, off-centred, acceptor defect in Figure 2.32(c). The acceptor-donor terminology is used in the literature in analogy to electronic band structures, and corresponds to the following meaning in photonic band gaps. A donor mode is generated by adding dielectric material to the structure, and results in a defect mode that is closer to the upper edge (high frequency end) of the stop band. An acceptor mode is generated by removing dielectric material from the structure, and results in a mode that is closer to the lower band edge. In the case of small defects, the donor or acceptor is "shallow" and occurs closer to the band edge. In the case of larger defects, the donor or acceptor is "deep" and is located further towards the centre of the stop band.

It may be seen in Figure 2.32 that an exceptional 50 dB attenuation has been recorded in the stop band. This is attributed both to the large dynamic range of the network analyser (HP 8510) and the sizable imaginary wave vector of the Yablonovite crystal. The maximum transmission of the defect modes is around -20 dB, which would not be particularly useful for a filter dealing with weak signals. Fortunately, transmission at the defect frequency may be improved by constructing the crystal from conductive materials instead of dielectrics.

2.6.6 Crystals having conductive materials

It is possible to construct photonic crystals from conductive substrates. This gives several advantages, including reduced loss (particularly for defect modes), a higher rejection rate in the stop bands, and more compact structures.

Metals are an excellent candidate for use as the conductive material at terahertz frequencies, and several metallic photonic crystals have been reported [123–126]. In the region of the plasma frequency and above (in the UV), dispersion and absorption effects must be taken into account when analysing metal structures. However, terahertz frequencies are several orders of magnitude lower than the plasma frequency and

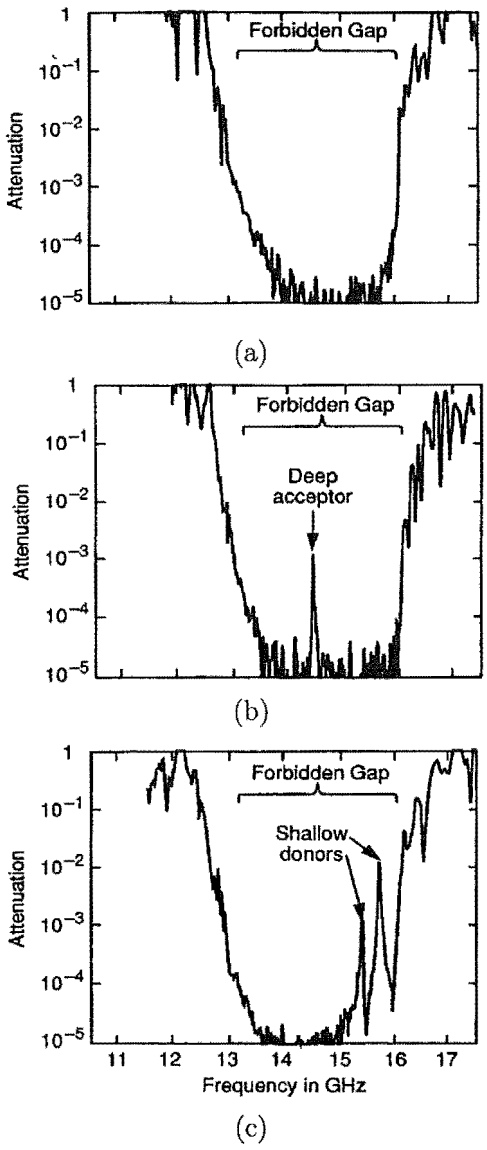


Figure 2.32 Transmission amplitudes obtained experimentally by Yablonovitch *et al.* in the microwave regime for a Yablonovite structure having a forbidden gap between 13 and 16 GHz [120]. (a) no defect, (b) single acceptor defect in the centre of the structure, (c) single donor defect off centre. It appears from the description in the text accompanying this figure in the original paper, that perhaps the vertical axis should be labelled "transmission coefficient".

therefore these effects may be neglected, allowing the metals to be treated as being essentially lossless. They may be accurately modelled as perfect electrical conductors. Other metal structures (not photonic crystals), have been known in the infrared for over three decades [7,127], providing additional validation of the suitability of metals for use at terahertz frequencies, since terahertz frequencies are at and below infrared frequencies. Alternatively, to take advantage of available microfabrication techniques, it is possible to use highly-doped semiconductors instead of metals, since the doped wafers can have resistivities as low as 0.002Ω-cm [128]. Structures using such low resistivity wafers are sometimes called "quasi-metallic" photonic crystals.

The same characteristic band gaps are observed with conductive photonic crystals as with dielectric photonic crystals, with two exceptions. Firstly, a continuously connected conductive structure (such as a woodpile) is impenetrable to long wavelengths. Propagation below a certain cut-off frequency f_c is prohibited

$$f_c \sim \frac{c}{2\Lambda}, \tag{2.35}$$

where c is the speed of light, and Λ is the period of the crystal. The long wavelength cut-off is is due to a plasmon-like resonance that has been investigated in some detail (see references in [124] for further information). This is different to dielectric structures, that appear as artificial dielectrics (having homogeneous but possibly anisotropic properties) and thus still transmit a significant portion of the incident wave, dependant on the average refractive index. Secondly, just above the cutoff frequency there is a transmission band that has (N -1) transmission resonances, where N is the number of layers in the crystal. The resonances arise from the formation of resonant microcavities between the layers.

A useful example of the behaviour of a metallic photonic crystal is given in Figure 2.33. The structure is a 2-D crystal comprising seven rows of metallic rods 2mm in diameter, on a lattice having period 6mm. This corresponds to a fill factor of 8.7% which is low compared to that of the 3-D dielectric crystals in section 2.6.4 ($\sim 30 \sim 80\%$). The first transmission band has six distinct resonances, equal in number to the six cavities that may be seen between the seven rows of rods. Characteristic of a metallic photonic crystal, there is no transmission for any frequency below the first transmission band.

To date, the highest reported transmission of a defect mode in a metallic photonic crystal is -2dB. As mentioned in the previous section, coupling defects to the outside of the crystal is important to achieve good transmission. However, if the defects are too closely coupled to each other, then the Q may be reduced or multiple defect modes (transmission modes) may appear. It appears that the level of coupling must be optimised to give the necessary Q and transmission.

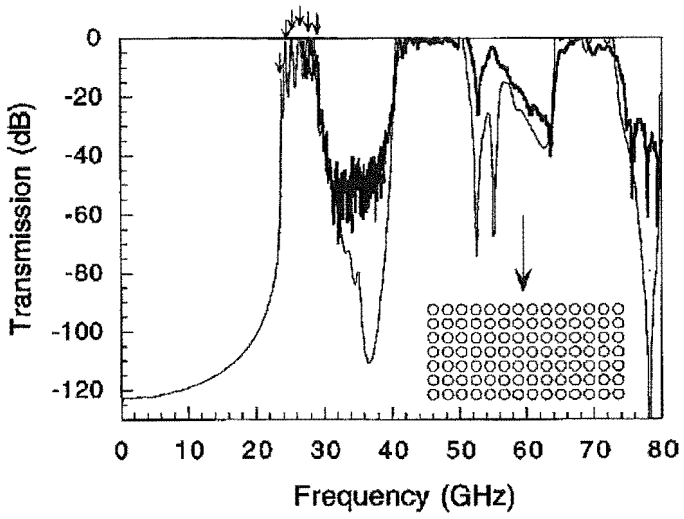


Figure 2.33 A 2-D metallic photonic crystal having seven rows of 2mm diameter metallic rods (structure shown in insert, large arrow indicates direction of propagation). The thin curve is a FDTD calculation, while the thick curve is a microwave measurement. Small arrows at the top of the graph point to transmission resonances in the first transmission band. Characteristic of a metallic photonic crystal, there is no transmission for any frequency below the first transmission band. Reproduced from [124].

2.6.7 Variable crystals

At the design stage, photonic crystals may be tailored with a great degree of freedom. However, it is more difficult to vary their properties "on the fly" after they have been fabricated. Several novel schemes have been proposed [129–132], involving either the physical manipulation of the crystal elements [129], or the alteration of the optical properties such as refractive index [130, 131] or free-carrier absorption [132]. Several examples are now presented, although it is by no means a complete list since it is believed that there are several studies presently underway, yet to publish details of their schemes.

An elegant scheme for the control of a 2-D crystal has been studied theoretically [129]. In this scheme, a hexagonal arrangement of air holes in a silicon substrate is mounted on a piezoelectric, magnetostrictive or electrostrictive substrate. When the crystal is subjected to a 2 - 3% shear strain from the actuation layer, the bandgap is predicted to shift by 52 - 73%.

The use of (nematic) liquid crystals as an alternative filling material to air has been variously investigated [130], and patented [131]. By altering the refractive index of the liquid crystal, the refractive index contrast in the photonic crystal may be altered. If the index contrast is sufficiently reduced, then the photonic band gap may be extinguished. Thus an electrically controllable switch or modulator may be produced.

Two differing methods of control were reported in the paper of Lourtioz *et al.* [132].

In the first method, diodes are inserted into the rods of a metallic 2-D photonic crystal. By electrically switching the diodes, defects may be created at will. Due to the cost of commercially-available varactors, the the integrated diode-controlled structure was studied by simulation only. In the second method, a laser is used to selectively illuminate added dielectric defects. The infra-red laser illumination induced free carrier absorption in the defects, extinguishing a small transmission peak in the stop band. One of the summary figures from this paper is reproduced in Figure 2.34. The modest amplitude of the transmission peak (~ -30dB, at 253GHz) is due not only to weak coupling between the incident beam and the defect, but also the modest resolution of the wide band terahertz spectroscopic system employed.

The transmission spectra reproduced in this section may be compared against the non-photonic-crystal tunable filters presented in the next section.

2.6.8 Other tunable filters

In this section, demonstrations of four alternative designs of tunable filter are presented, the first three of which are based around varying some aspect of a multilayer stack. While extensive use was made of the multi-layer stack as a 1-D crystal example in sections 2.6.3 and 2.6.5, the three multilayer designs presented in this section may not be considered as photonic crystals designs, because they have been derived according to mature design principles from visible optics.

The first filter is a tilting multilayer stack in a gas spectrometer [81], as shown in Figure 2.35, with a diagram of the spectroscopic system in Figure 2.35(a) and a plot of transmission spectra at several different angles of tilt in Fig. 2.35(b). The centre wavelength of transmission or reflection decreases with increasing angle of incidence according to Bragg's law

$$\lambda_\theta = \lambda_0 \sqrt{1 - \left(\frac{\sin\theta}{n}\right)^2} \tag{2.36}$$

where λ_θ is the centre wavelength at angle of incidence θ, while λ_0 is the centre wavelength at normal incidence, and n is the effective index of refraction for the whole stack.

The tunable Fabry-Perot filter of Bondavelli *et al.* is designed for the visible, (optical communications in particular), and uses electrostatic force to deform the two InP layers either side of a cavity, shortening the effective cavity length and shifting the passband up in frequency [133]. The filter is shown in Figure 2.36, with the structure in Figure 2.36(a) and the optical reflectivity spectra in Figure 2.36(b). The reflective mirrors either side of the cavity are made from quarter-wave layers of InP and air. Tuning voltages do not exceed 15V.

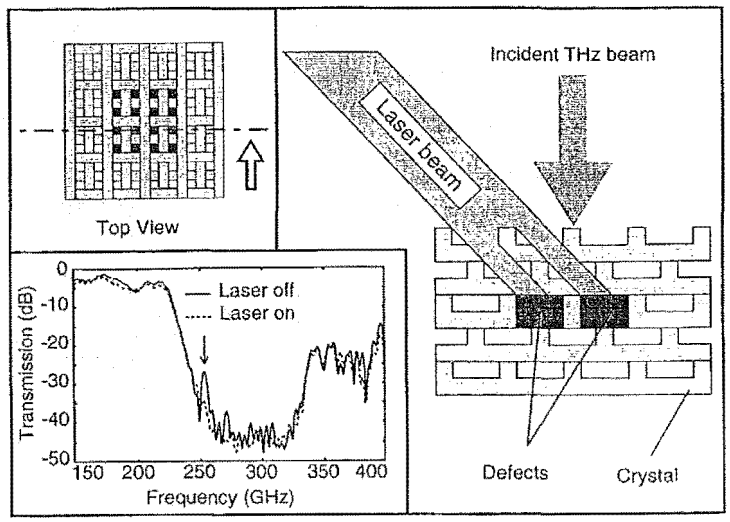


Figure 2.34 Lourtioz's scheme for optically altering the level of free carrier absorption in buried defects. The top left picture shows a plan view of the woodpile-like crystal and the location of the buried defects. The right picture shows the laser beam incident from an off-normal angle, while the THz beam is incident on the normal. The measured millimetre wave spectra is plotted in the lower left panel. The solid curve is with the laser off. The dashed curve is with the laser on. Turning on the laser extinguishes the defect mode at 253GHz (indicated by the arrow).

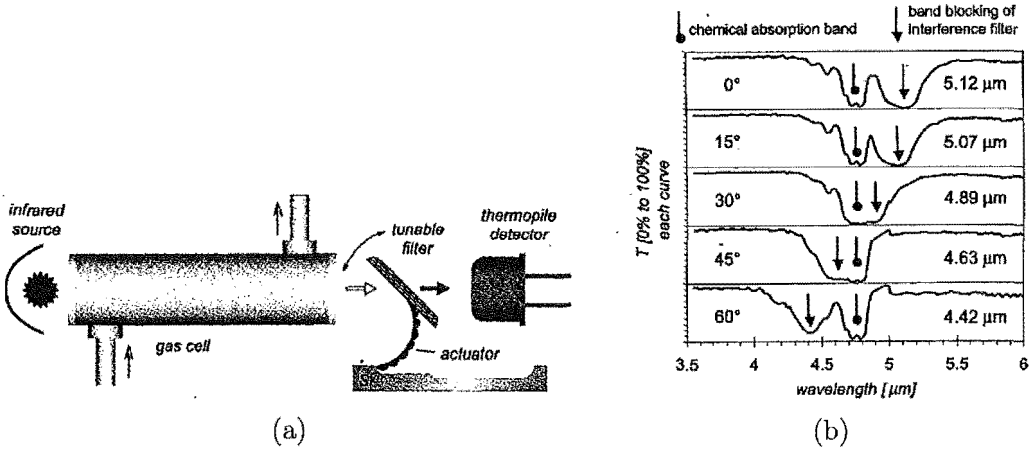


Figure 2.35 The MEMS-tunable tilted multilayer stack interference filter of Lammel *et al.* [81]

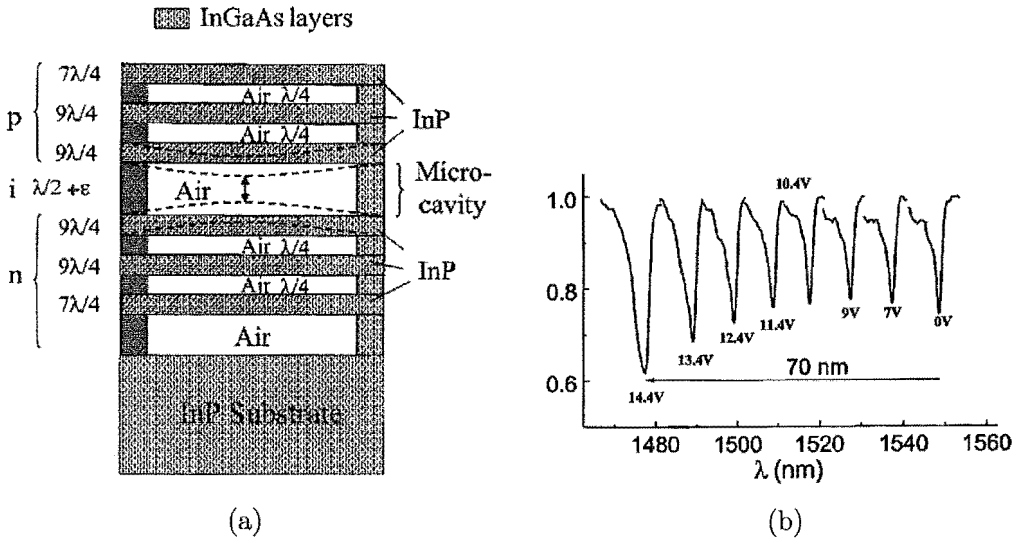


Figure 2.36 The electrostatically-tunable cavity Fabry-Perot filter of Bondavelli *et al.* [133]. (a) The Indium-Phosphide - air multilayer structure and tunable cavity, and (b) optical reflectivity spectra.

The gas sensing filter of Alause *et al.* is targeted to infrared wavelengths between 2 and 8 μm [134]. The filter is presented in Figure 2.37, with the filter structure and method of tuning in Figure 2.37(a), and the transmission spectra for three representative settings in Figure 2.37(b). Like Bondavelli's structure, a tuning voltage (< 20V in this case) is applied to the structure, shifting the central wafer off-centre and lowering the centre frequency of the passband.

Gianvittorio *et al.* investigated a novel form of a frequency selective surface, comprised of tiltable metallic dipoles [135]. The filter is presented in Figure 2.38, with the filter structure and supporting dielectric substrate shown in Figure 2.38(a), and the transmission spectra for several representative settings in Figure 2.38(b). Increasing the tilt angle away from the plane of the substrate shifts the stop band up in frequency, but unfortunately reduces the rejection.

This section has presented four types of tunable filter. At least two of the filters have a tuning range in excess of their bandwidth. The filter with the narrowest band (highest Q) is that of Bondavelli *et al.*, and it exhibits an impressive tuning range of at least eight times its bandwidth. Unfortunately the rejection rate is low, about 4dB at best (bias = 14.4V).The filter of Gianvittorio *et al.* had much better rejection (up to 18dB), but the tuning range was approximately equal to the bandwidth and the rejection reduced to 6dB in the highest frequency stop band. The filter of Alause *et al.* can tune at least twice the bandwidth and has rejection of about 13dB (assuming the bottom of the graph corresponds to zero). The high defect transmission and strong rejection of a metallic photonic crystal compare favourably to these values, and the photonic crystal structure would therefore represent an improvement in performance if

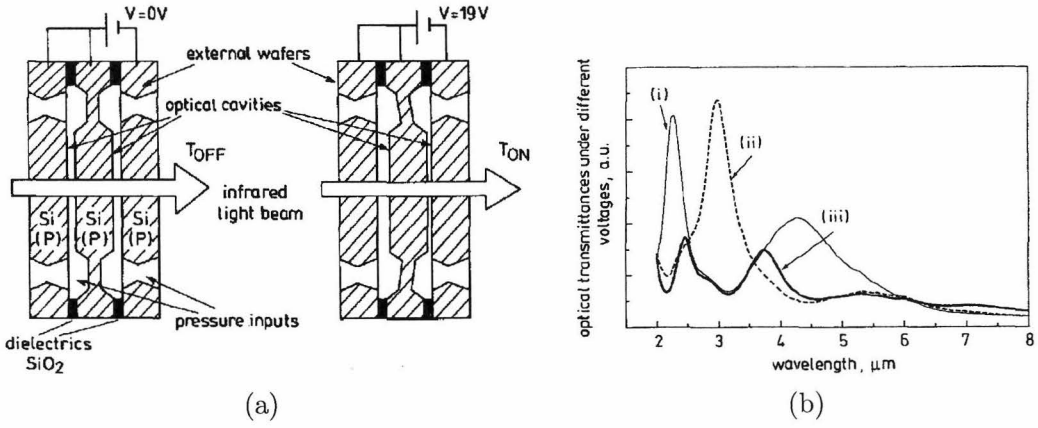


Figure 2.37 The tunable dual-cavity filter of Alause *et al.* [134], (a) structure in "on" and "off" positions, (b) transmittance as a function of applied bias voltage, (i) = 0V, (ii) = 18V, (iii) = 19V.

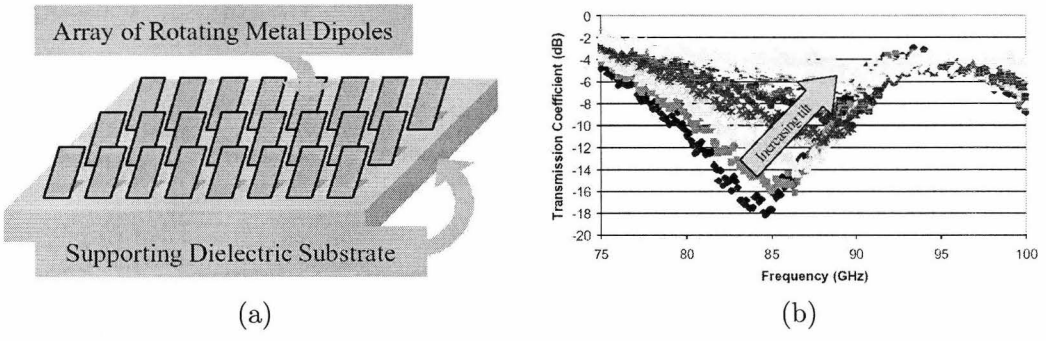


Figure 2.38 The reconfigurable frequency selective surface, tunable filter of Gianvittorio *et al.* [135]

a large enough tuning range could be achieved.

2.6.9 VADR-inspired variable PBG filter

The key concept behind VADR is that the degree of separation between two interlocking sub-wavelength gratings may be varied to control the retardance. In this study, inspired by the concept of VADR, a similar method was used to create a tunable filter for frequency selective filtering.

The concept is illustrated in Fig. 2.39. The basic structural element is a plate having two perpendicular layers of rods (not shown), and the separation between any or all of these plates may be altered at will to give an unprecedented level of control over the number and type of defects in the photonic crystal. A device with all the plates fully interlocked is shown in Fig. 2.39(a), while a device with the plates only partially interlocked is shown in Fig. 2.39(b). The unit cell of the crystal is shown in Fig. 2.40.

Nothing like this concept for the "on-the-fly" tuning of a 3-D periodic structure has been presented in the literature before. Other tuning schemes that have been presented are completely different. For example, one prominant example involves the infiltration of a fixed structure with liquid-nematic crystals and then electrically altering the refractive index of the liquid crystals to modulate the refractive index contrast above and below the critical threshold for the formation of a bandgap ($n_s/n_i \geq 1.9$) [114].

2.7 SUMMARY

This chapter has provided detailed coverage of background material relating to the two devices investigated in this thesis, the variable retarder (a polarisation control device), and the tunable filter.

It has long been known that naturally-occuring birefringent crystals are able to transform the polarisation of visible light. A great many devices have been designed along these lines, including waveplates having fixed retardance, and compensators having variable retardance. Unfortunately, at terahertz frequencies, the optical properties of these crystals deteriorate to the point that the devices become practically useless.

Much better devices can be made from silicon, at a fraction of the cost, since it has excellent optical properties at terahertz frequencies and may be readily made birefringent with artificial dielectric techniques. The concepts of artificial dielectric waveplates, and uniaxial-crystal compensators may be combined to arrive at a variable artificial dielectric retarder. Such devices have never before been demonstrated, and the work in this thesis is largely devoted to the fabrication, measurement, analysis and improvement thereof.

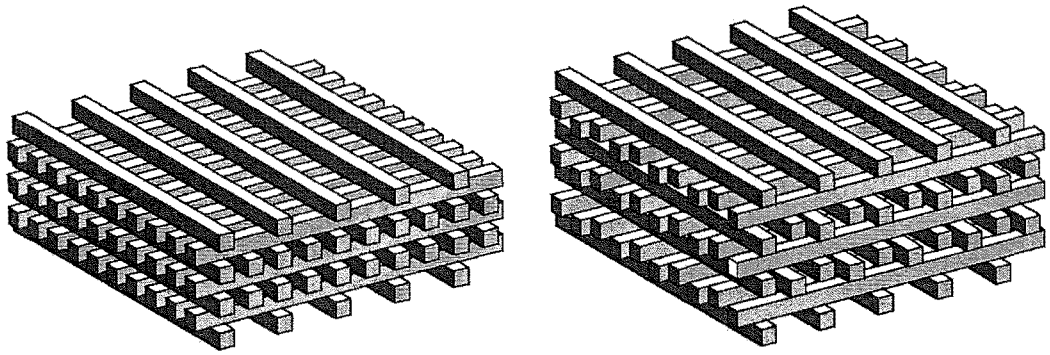


Figure 2.39 The variable photonic band gap filter concept: plates with two perpendicular layers of rods interlock to form a layer-by-layer photonic crystal with unprecendented control over the number and type of defects. (a) all layers fully interlocked, (b) all layers partially interlocked.

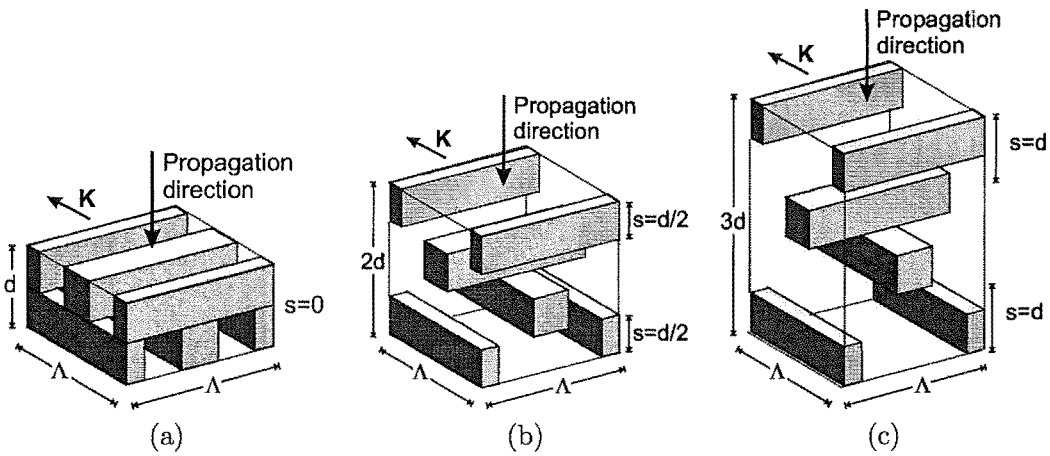


Figure 2.40 Unit cell of the variable photonic band gap filter device. Shaded areas represent the substrate. The propagation direction is indicated, as are the grating vector **K**, rod period Λ, plate depth d (rod depth is $d/2$) and plate separation s. (a) fully interlocked ($s = 0$), (b) just interlocking ($s = d/2$), (c) widely separated ($s = d$).

One of the improved variable retarders developed in this work inspired a novel method for the "on-the-fly" tuning of photonic crystal filters. When the periodic structure of a photonic crystal is disturbed, it often leads to the presence of a narrow pass band within the broad, characteristic, stop band. By dynamically controlling the position of the layers of a photonic crystal in a number of different ways, a variety of tunable filter effects may be achieved, one of which is demonstrated experimentally in this work.

Chapter 3

SILICON MICROFABRICATION TECHNIQUES

3.1 INTRODUCTION

This chapter gives a background to the micromachining techniques used in the fabrication of the VADR plates. While it describes the techniques, equipment, and process recipes in approximately the order in which they were used, the details of the VADR design and overall fabrication procedure are left until the next chapter. The description begins with the wafer scriber, used to mark the wafers before cleaving them into suitably-sized samples. It continues with contact printing photolithography, used to pattern photoresist on silicon nitride (Si_3N_4)-coated silicon substrates. The Si_3N_4 was then reactive ion etched to become a wet etch mask. The mechanisms of RIE etch process are described along with the principles of operation of the RIE machine, and the process recipe. The alkaline wet etching of silicon is a fundamental micromachining technique, used here to create the long, deep V-grooves of the VADR plates. Important aspects include the nature of silicon lattice, identification of crystal planes, and the geometries that result when those planes etch at different rates. The wet etch equipment is described, along with a process recipe that was formulated to eliminate the formation of hillocks. A computer-based etch simulation technique was used to verify some mask designs before committing them to production; the technique is briefly described and an example of the program's output is provided. Except where otherwise stated, existing University of Canterbury (UoC) laboratory procedures are described in this chapter.

3.2 FABRICATION OVERVIEW

This section provides a short overview of the VADR fabrication process, so that the following process descriptions may be placed into context. The fabrication steps are illustrated in Fig. 3.1. The silicon wafers are pre-coated in silicon nitride (Si_3N_4), as shown in Fig. 3.1(a). The Si_3N_4 is patterned using photolithography and reactive ion etching to produce a mask, as shown in Fig. 3.1(b). The silicon substrate is then etched

anisotropically to form V-grooves, as shown in Fig. 3.1(c). Finally, the Si_3N_4mask is removed to leave the V-grooves as shown in Fig. 3.1(d).

3.3 WAFER SCRIBING

In order to break a wafer cleanly into smaller square samples for experimental work, it must be inscribed along the lines where it is to be broken. For this, a diamond-tipped tool is of great use. The University of Canterbury (UoC) built their own scriber, pictured in Figure 3.2. A wafer, not shown, was normally mounted on the platform, and held in place by vacuum. The stylus was positioned laterally by a calibrated screw thread, marked "Positioning Control" in Figure 3.2. To mark a line, the stylus was brought into contact with the wafer by releasing the handle marked "height control". Once on the wafer, pressure was provided by a spring arrangement. The screws marked "pressure control" were pre-adjusted so that a firm, even pressure was applied by the stylus throughout each stroke. Each line was marked in a single stroke, by manually sliding the platform back along its single-axis track. Once the first set of lines was marked, the top half of platform was able to be rotated by exactly 90° so that the second, perpendicular, set of lines can be scribed.

Since a silicon wafer only breaks cleanly along certain crystal planes (such as {111}), it was important to mark the scribed lines parallel to the wafer flats. To align the wafer, it was placed in the platform without applying a vacuum. A microscope with a micrometer graticle in the eye piece was positioned over the primary wafer flat, to provide a positional reference. The platform was moved back and forth along its single axis until the edge of the primary flat remained within one major division of the graticle along its entire length. Since the wafer flat was approximately 30mm long, and one major graticule division corresponded to 0.5mm, the scribe marks were aligned to within 1° of the wafer flats.

3.4 PHOTOLITHOGRAPHY

In the context of micromachining, photolithography is the process of transferring a two-dimensional (2-D) pattern or drawing to a layer of photosensitive material on the surface of a substrate. After photolithography, the 2-D pattern can be converted to a 3-D structure by additive (deposition) or subtractive (etching) processes, or exposed areas can be doped, as illustrated in Figure 3.3. There are a number of methods of photolithography, and most have been developed with the aim of patterning features of ever-decreasing dimensions. The method of contact printing was adopted because it was the standard photolithographic procedure in the UoC laboratory. Contact printing was one of the earliest methods adopted by the semiconductor industry and it was used extensively for the patterning of 3–10μm images across entire wafers [136], and it is

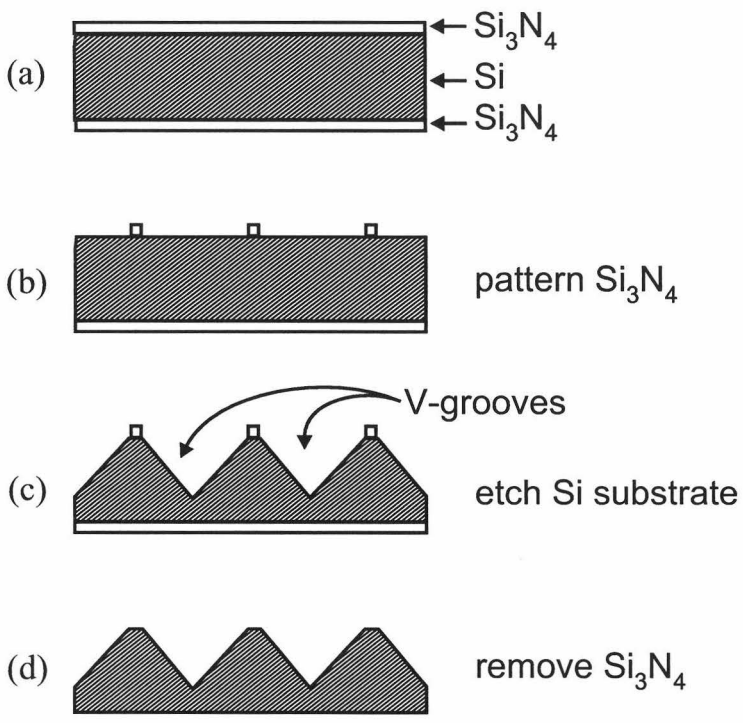


Figure 3.1 Overview of VADR fabrication process. (a) Silicon wafer pre-coated with Si_3N_4, (b) pattern Si_3N_4, (c) anisotropic etch of silicon substrate, (e) remove Si_3N_4.

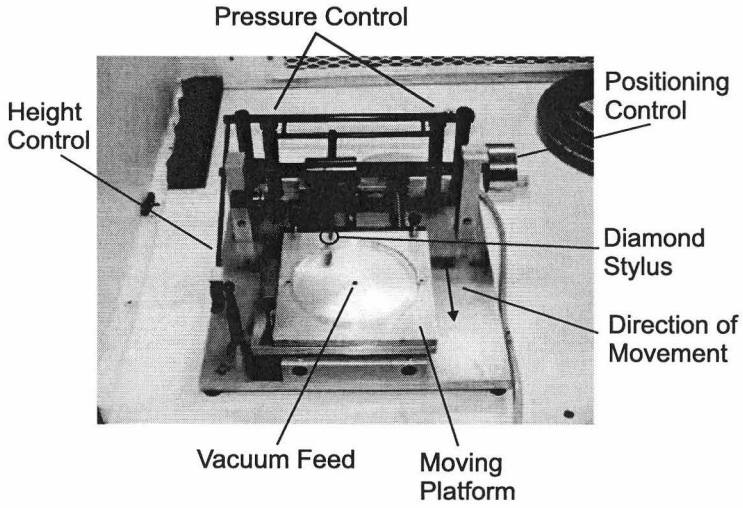


Figure 3.2 The home-made wafer scriber in the UoC laboratory. A diamond-tipped stylus inscribes the wafer surface, allowing the wafer to be cleanly cleaved.

certainly capable of reproducing the 5–40μm features required by the VADR fabrication task.

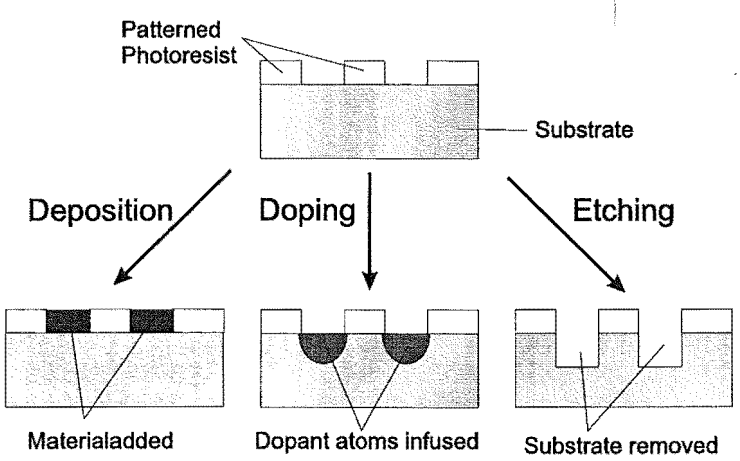


Figure 3.3 The planar fabrication processes, deposition (a), doping (b), etching (c).

The procedure for photolithography with S1813 positive photoresist [137], on 25mm by 25mm samples of silicon wafer, was as follows. Initially, each sample was cleaned by immersion and flowing rinse in acetone, iso-propyl alcohol and then methanol. They were then blown dry with nitrogen, and dehydrated in an oven at 95°C for 15 min. A home-made photoresist spinner was used, and it is pictured in Figure 3.4. One sample at a time was placed on the photoresist spinner's chuck, and held in place with a vacuum. The spinner was set to operate at 4000rpm for 60s, including 1-2s for acceleration at the start and 1-2s for deceleration at the end. Both the position of the sample on the chuck (it had to be centred), and the suction grip, were checked with a trial spin without photoresist, during which the spin speed was also fine tuned. The sample was brought to rest after the trial spin. Two or three drops of photoresist were expelled from a 1ml disposable pipette onto the centre of the sample, to give a dispense volume in the order of 100μl. The volume contained in a 1μm layer of photoresist on a 25mm x 25mm sample is 0.63μl, so much of the photoresist was lost to waste during the spin, however the large initial quantity was necessary to ensure sufficient inertia in the photoresist to overcome surface tension and coat the entire wafer surface. The sample was spun for the preset 60s. After spinning, the sample was soft-baked at 95°C for 25 min. If the sample was baked at a higher temperature or a longer time, it was difficult to dissolve the exposed photoresist areas during development. After soft-bake the sample was allowed to cool to room temperature, while the mask aligner (see below) was set up. The sample was exposed for 45s, then developed by immersion in undiluted MF320 developer for 35s, followed immediately by a rinse in DI H_2O. Microposit MF320 is an alkaline developer containing tetra methyl ammonium

hydroxide (TMAH). The sample was blown dry with N_2, and hard-baked at 115°C for 25 min to help the photoresist withstand the reactive ion etching. No experiments were carried out to quantify the effects of hard-baking on the photoresist etch rate because the standard procedure routinely gave acceptable results.

A negative photoresist was also used at times. The particular product, AZ5206 [138], was an image reversal positive photoresist using the same Novolak resin as S1813, except that it has been modified so that the positive image formed during the initial exposure can be reversed to a negative image by baking off CO_2 and flood exposing (exposing with no mask). A negative-acting photoresist is ideal for use with the wagon wheel alignment mark (described in detail in section 4.4), that has only a small number of mask openings, because it considerably reduces the time required to photoplot the mask reticle. The procedure for photolithography with AZ5206 was as follows. The sample was cleaned and dehydrated as it was for use with S1813. The photoresist spin procedure was also the same, except that 6-8 drops of photoresist were put on the sample and the final thickness of the AZ5206 was 0.6μm. The sample was soft-baked on a hot-plate for 90s at 90°C, cooled, and exposed for 6s with the mask. Then it was reverse baked at 120°C for 90s, cooled, and flood exposed (no mask) for 20s. Development took 60s in AZ300MIF, that, like MF320, contains TMAH. The photoresist was hardened by a further bake, at 115°C for 1 min. This was half the manufacturer's recommended time, to make it easier to clean off after subsequent processing, such as reactive ion etching.

3.5 REACTIVE ION ETCHING

Reactive ion etching (RIE), also known as dry etching, is used for etching the Si_3N_4 coating on the VADR silicon samples. Dry etching has the advantage of offering precise control over etch conditions and side wall profile. There are also fewer waste products to dispose of, and the technique is safer from the operator's point of view than the alternative technique of wet etching with hydrofluoric acid (HF) acid. Dry etching is also more convenient because a photoresist mask is sufficient, whereas for wet HF etching, extra processing steps are required to pattern a nickel-chromium mask. The HF technique suffers the additional drawback of requiring the backside coating of Si_3N_4 to be protected.

3.5.1 Equipment Description

The UoC laboratory has an Oxford Plasmalab System 80 Plus, configured for single chamber RIE operation, represented in the schematic of Figure 3.5. The machine has four basic subsystems - the chamber where the etching takes place, the process gas

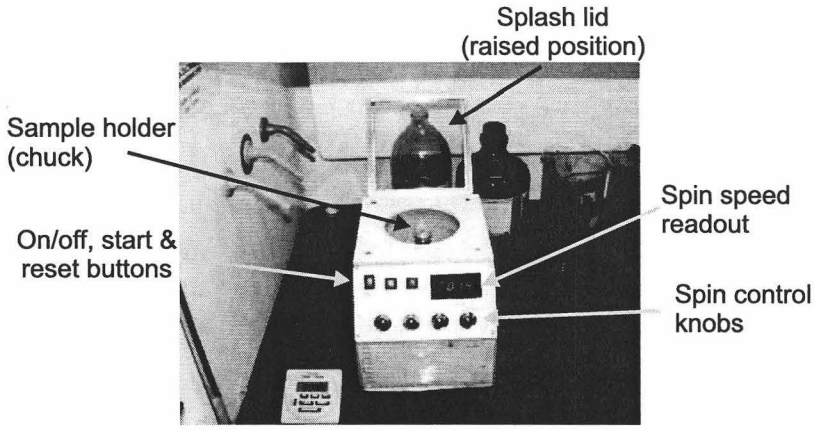


Figure 3.4 The UoC laboratory's home-made photoresist spinner

handling system, exhaust gas and vacuum pumping system and the microprocessor control system.

The chamber is configured with two parallel electrodes. The bottom electrode is 200mm in diameter, and is connected to a variable 200W RF power supply that operates at a frequency of 13.56MHz. The bottom electrode can be cryogenically cooled with liquid N_2 to temperatures as low as -150°C, and heated with an internal element to 200°C. The process gas inlet is at the top of the chamber, and the waste gas exhaust is at the bottom. The inside top surface of the chamber is the grounded top electrode. The electrodes alternate as anode and cathode throughout the etch due to the inherent AC nature of the RF power supply. The etch mechanism is described in detail in section 3.5.2.

The available process gases, SF_6, O_2, CHF_3, Ar, N_2, and CH_4, are multiplexed into three input gas lines, as shown in Figure 3.5, allowing a maximum of three gases into the chamber during any one etch. Individual mass flow controllers (MFC) are used to regulate the gas flow rate in each of the three lines. Downstream of the MFCs, the three lines are combined into one line, for entry into the chamber. A separate line is supplied for the vent gas, which here is N_2.

Vacuum is provided by a combination of a turbomolecular pump and a rotary roughing pump. The rotary pump is used to reduce the pressure from atmospheric until the turbo pump is able to take over. The chamber must be reduced to a base pressure of $1x10^{-7}$Torr before etching can begin. The rotary pump also backs the turbo pump, isolating it from atmospheric pressure. The exhaust gases are routed to the rotary pump, via a the automatic pressure control (APC) valve, which is operated by the microprocessor control system to maintain the chamber pressure to a preset level.

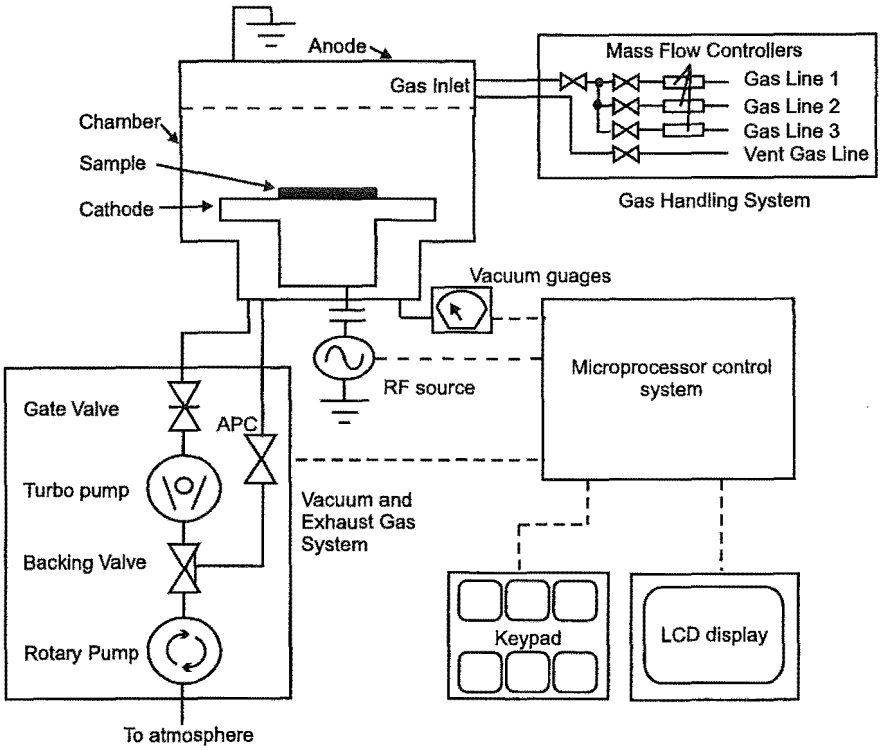


Figure 3.5 Schematic diagram of Oxford Plasmalab 80+ Reactive Ion Etcher

The microprocessor control unit can be operated from the local keypad and LCD display or via a remote PC terminal using Oxford's PCPLUS software in a Microsoft Windows environment. The local interface system uses a series on menus to enter process parameters such as temperature, pressure, gases, gas flow rates, etch duration and power. During the etch, the control unit monitors the process conditions and displays them on the LCD screen, and controls the APC and MFCs.

3.5.2 Dry Etching Si_3N_4

The RIE machine was used to anisotropically etch 200nm films of chemical vapour deposited (CVD) Si_3N_4, on a silicon substrate. The patterned Si_3N_4 was used as an etch mask for the wet anisotropic etching stages of the VADR plate fabrication process. The etch requirements are not stringent as they may be in other applications, because the smallest mask feature size is 5μm, which is the width of the "spokes" in the dual radial alignment mark of section 4.4.3. With a Si_3N_4 layer depth of 200nm, even an isotropic etch will provide suitable line width control, as 400nm extra width in any of the openings will not adversely affect the final etch results.

A CHF_3 and Ar based etch chemistry has been shown to be ideal for this procedure [139]. In the plasma, the CHF_3 provide F atoms which are responsible for the chemical etching of Si_3N_4. Ar is introduced to physically bombard the etch site to remove etch-inhibiting polymer deposition and to open up new sites for the F to attack the Si_3N_4. CHF_3 provides Fluorine atoms via Equation (3.1) that chemically etch the Si_3N_4 to produce tetrafluorosilane (SiF_4) and nitrogen via Equation (3.2).

$$e^- + CHF_3 \rightarrow CHF_2^+ + F + 2e^- \tag{3.1}$$

$$Si_3N_4 + 12F \rightarrow 3SiF_4 + 2N_2 \tag{3.2}$$

A CHF_3 and O_2 etch is similarly effective, however it has the inconvenient consequence of requiring a nickel chromium etch mask, because photoresist is readily attacked by oxygen, destroying it before the etch is complete.

An anisotropic process with a good etch rate had already been developed in the UoC laboratory for the successful creation of Si_3N_4 wet etch masks [140], so this procedure was adopted, although the author made two minor modifications for the sake of expediency. The etch conditions for the modified process are presented in Table 3.1. The first modification was to the etch temperature. The original process called for cryogenic etch at -20°C. It was determined by experiment that the process could also be used at room temperature, halving the total process time by eliminating time-consuming temperature changes in the chamber. The self-bias voltage, V_b, of the room temperature etch was approximately -470V.

Table 3.1 RIE etch parameters

Gases	Etch Rate (nm/min)	Pressure (mTorr)	Flow rate (sccm)	Power Density (W/cm^2)	Temperature (K)
CHF_3/Ar	25	25	50/35	0.63	295

The second modification was to the masking material. In the first few rounds of dry etching trials, a NiCr metal etch mask was used. The NiCr mask was patterned by forming a negative image in photoresist, evaporating 20nm of NiCr over the top, and then immersing the sample in acetone, which dissolved the photoresist, and lifted-off the unwanted areas of the NiCr mask. Approximately one day of lab time per sample-set would be eliminated if the NiCr mask could be done away with in favour of using photoresist as an etch mask. It was determined by experiment that a 1μm layer of S1813 photoresist was a suitable etch mask. The etch rate of the photoresist was not quantified, as there was a sufficiently thick layer remaining after etching, that no further optimisation of the RIE process was required to obtain good results routinely. The samples were immersed in a KOH etch solution to check that the dry etch had gone to completion, since silicon etches (and bubbles are formed) when exposed to a KOH etch solution but Si_3N_4 does not. Upon immersion, bubbles were immediately observed in the area where the Si_3N_4 had been etched to expose the silicon substrate, but not where it was masked, indicating that the RIE etch had gone to completion.

3.6 WET ETCHING OF Si_3N_4

Procedures exist for the wet etching of Si_3N_4 in solutions of either hydrofluoric or phosphoric acid. For example, a concentrated solution of 40% wt/wt HF in water etches isotropically at a rate of 133Å/min [140]. However, masks made of photoresists, such as S1813, are found to peel off within 5-6 min of immersion in 40% wt/wt HF, whereas it takes 15 min to etch through the 200nm coating of Si_3N_4. The HF does not attack the photoresist directly, but rather severely reduces its ability to adhere to the Si_3N_4 surface. A NiCr mask is a suitable alternative as it does not suffer from reduced adhesion when immersed in HF, however it does etch at $\frac{1}{6}$ of the rate of Si_3N_4. In this case, an underetch of 33nm would be suffered, but this is of little consequence with a minimum feature size of 5μm, and a process tolerance in mask dimensions of at least 1μm.

It has been reported that buffered HF and diluted HF do not attack the adhesion between the photoresist and the substrate [141], allowing photoresist to be used as a mask. However, the Si_3N_4 etch rate falls by a factor of at least ten in these solutions. The reported etch rates are 11 Å/min for 1:10 HF:H_2O by weight, and 9 Å/min for 1:5 buffered HF, giving etch times of at least 3 hours. Phosphoric acid can also be

used to etch Si_3N_4, although it must be heated to 160°C to raise the etch rate to a modest 30Å/min. For safety reasons, it is desirable to avoid high temperature etching when room temperature alternatives exist. Photoresist is not suitable as a mask because phosphoric acid, like HF, reduces the adhesion of the photoresist. In order to use a NiCr mask for patterning, additional processing steps must be undertaken. After photoresist pattering of a negative mask, NiCr is evaporated onto the front surface. A lift-off procedure, such as immersion in acetone and ultrasonic agitation, removes unwanted metal. The Si_3N_4 coating on the backside of the sample must also be protected with a layer of NiCr, and this requires an entire second evaporation process unless the sample can be turned in the chamber.

These wet etch techniques are of most use in the complete removal of the Si_3N_4 coatings after the V-grooves have been formed, and are of little use in patterning due to the slow etch rates in solutions that do not attack photoresist masks.

3.7 WET ETCHING OF SILICON

The most important part of the VADR plate fabrication process is the anisotropic alkaline wet etch that creates the V-grooves and leaves them with a mirror smooth finish. It is an ideal technique for fabricating the VADR plates, for while it is possible to fabricate grooves of a variety of cross-sections using dry-etch techniques, the deep V-grooves required would be extremely difficult to produce with the particular reactive ion etching (RIE) machine available in the UoC laboratory. Additionally, the wet etch equipment and consumables are considerably less expensive.

The properties of crystalline silicon are highly anisotropic, and this can be exploited to create complicated geometries with relative ease through the use of orientation-dependent etch solutions. The orientation of the crystal lattice with respect to the wafer surface is important in these reactions, as this determines the type and geometry of the features that can be anisotropically etched. It is useful to review the lattice structure of silicon and the Miller indices system of describing crystal planes within that lattice, before describing the features that may be achieved easily using silicon wafers of the two most widely available crystal orientations. The mechanisms of mask undercutting during etching are explained. The etching equipment is described, along with measures to ensure consistent conditions over long etches. Hillocks are an unwanted etch feature, characteristic of certain etch conditions. Their formation and a etchant formulation that prevents it, are described. This etch formulation is used in the fabrication of the VADR plates.

3.7.1 Crystalline silicon

The atomic nuclei of a crystalline solid are arranged in a lattice structure that is periodic in all three dimensions. The whole lattice can be described by a small unit cell, that when repeated in all three directions, recreates the entire lattice. The silicon lattice is shown in Figure 3.6. It consists of two face centred cubic unit cells, with the second cell offset from the first by $(\frac{1}{4}a, \frac{1}{4}a, \frac{1}{4}a)$. This arrangement is also found in diamond, giving it the name diamond lattice. The lattice constant, a is an important parameter because it is a measure of the size of the unit cell. The lattice constant is equal to the length of the side of the unit cell, and in the case of silicon at room temperature is 5.430Å. Silicon atoms bond in a tetrahedral arrangement, with 109.5° angle between the bonds, and bond length of 2.352Å.

3.7.2 Miller indices

The properties of crystalline silicon are anisotropic, therefore it is appropriate to distinguish between the various crystal planes in the silicon lattice. The commonly-used Miller indices define crystal planes in the unit cell by their normal vector. In order to generate the Miller index of a particular plane, the plane in question is drawn on one or more unit cells, and the intercepts with the x,y, and z axes are found, and expressed in terms of the lattice constant, a. The reciprocals of the three numbers are taken, and the results are reduced to the smallest three integers having the same ratio. The results are enclosed in parentheses, and overbars are used to indicate any negative numbers. The low-index planes (100), (110), (111) are the most useful for describing anisotropic etching. These planes are illustrated in Figure 3.7.

One of the most important characteristics of a silicon wafer is the plane to which the polished top surface has been aligned. The miller index of this plane is often used when describing the wafer. For example, a (100) silicon wafer has the top surface aligned with one of the (100) planes. Within the lattice, there are groupings of planes that are equivalent in atomic packing density, and these planes are often referred to together. For example, the (100) and $(\bar{1}00)$ planes are equivalent, and described as a group they are {100}, with curly brackets. (100) Silicon wafers have two flats, that indicate the directions of the (110) planes. The vector normal to the (110) plane is called the [110] direction, with square brackets. Because there are planes equivalent to (110), each with its own crystallographic direction, the corresponding group of crystal directions is called ⟨110⟩, with angle brackets. For reference, these conventions are summarised in the front matter of this thesis.

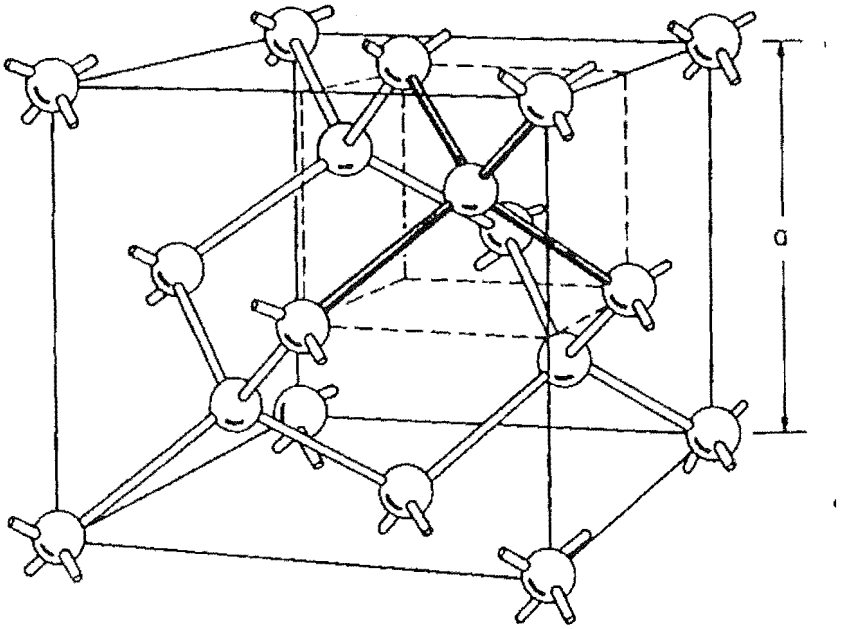


Figure 3.6 Unit cell of the silicon lattice. (Reproduced from [142] Figure 1.3)

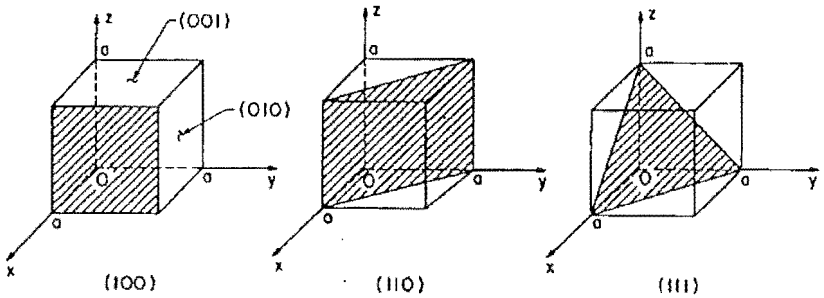


Figure 3.7 Low index crystal planes and their Miller indices. (Reproduced from [142] Figure 1.2)

3.7.3 Anisotropic Wet Etch Mechanisms

The etch behaviour of single-crystal silicon in anisotropic etchants is well known, and the technique is widely used. Anisotropic etching is also known as orientation dependent etching, because the solutions etch the silicon much faster in one direction than another. Typically, the {111} planes etch the slowest, and these are sometimes referred to as the "etch-stop" planes. The other planes of interest are {110} and {100}. Taking the etch rate of the {111} planes as a reference , i.e. {111} = 1, the values for relative etch rates may be as high as {100} = 400 [143], {110} = 600 [144]. These values are strongly dependent on the chemical composition, concentration, and temperature of the etchant, and are by no means typical of all solutions. It is this difference in etch rate that accounts for the ability of anisotropic etches to create geometries that would be difficult or impossible with other techniques.

There are several different classes of chemical that anisotropically etch silicon [145]. Firstly, there are the hydroxides of alkali metals such as potassium (KOH), sodium (NaOH), caesium (CsOH), Rubidium (RbOH), etc., that exhibit the high etch rate anisotropies. Secondly, there are the simple and quaternary ammonium hydroxides, such as ammonium hydroxide (NH_4OH) and TMAH($(CH_3)_4NOH$), which unfortunately produce greater surface roughness, and exhibit less etch rate anisotropy than the alkali metal hydroxides. Thirdly, there are two etches based on ethylene diamine - with pyrochatechol it is known as EDP, and with pyrochatechol and water it is known as EPW. These etches are hazardous, being both corrosive and carcinogenic, and give relative etch rates for the {100}:{111} planes in the order of 35:1, an anisotropy lower than that of either the ammonium or alkali metal hydroxides. Fourthly, there are other less popular etchants, such as hydrazine. Hydrazine has a useful etch rate but exhibits less anisotropy than KOH and EDP, and is equally hazardous as EDP.

For the fabrication of the VADR plates, a high etch rate anisotropy is preferred, to minimise mask undercut, and minimal surface roughness is important. Thus, the alkali metal hydroxides present the best option. Potassium hydroxide (KOH) was chosen because it was readily available, and its behaviour was well described in the literature. Despite there being a good understanding in the literature of how to exploit the KOH etchant, a description of its chemistry, and that of the other etchants, has yet to be finalised. The large anisotropy in the etch rate of the different planes has yet to be correlated to a property of those planes. For example, the surface density of atoms varies by only a few percent across all directions, and this, alone, simply cannot explain the large etch rate anisotropies that have been observed.

Currently, it is accepted that the reaction sequence is as follows [146]. Hydroxyl ions from the etchant react with the silicon atoms at the surface, oxidising them. The oxidation process promotes four electrons from each silicon atom in to the conduction

band

$$Si + 2OH^- \rightarrow Si\,(OH)_2^{2+} + 4e^-. \tag{3.3}$$

The water is reduced, producing hydrogen

$$4H_2O + 4e^- \rightarrow 4OH^- + 2H_2. \tag{3.4}$$

The formation of hydrogen bubbles can be clearly seen during the etch, and some surface roughness is attributed to the micro-masking effect of the bubbles. Further hydroxyl ions react with the oxidised silicon, $Si\,(OH)_2^{2+}$, to form a soluble silicon complex and water

$$Si\,(OH)_2^{2+} + 4OH^- \rightarrow SiO_2\,(OH)_2^{2-} + 2H_2O. \tag{3.5}$$

The overall reaction is thus

$$Si + 2OH^- + 2H_2O \rightarrow SiO_2\,(OH)_2^{2-} + 2H_2. \tag{3.6}$$

3.7.4 Wet Etch of (100) Silicon

(100)-silicon wafers have, by definition, {100} surface planes, and ⟨110⟩ as the wafer flats, as shown in Figure 3.8. Note that (100)-silicon wafers often only have one flat cut, unlike those used in this study that had two wafer flats cut.

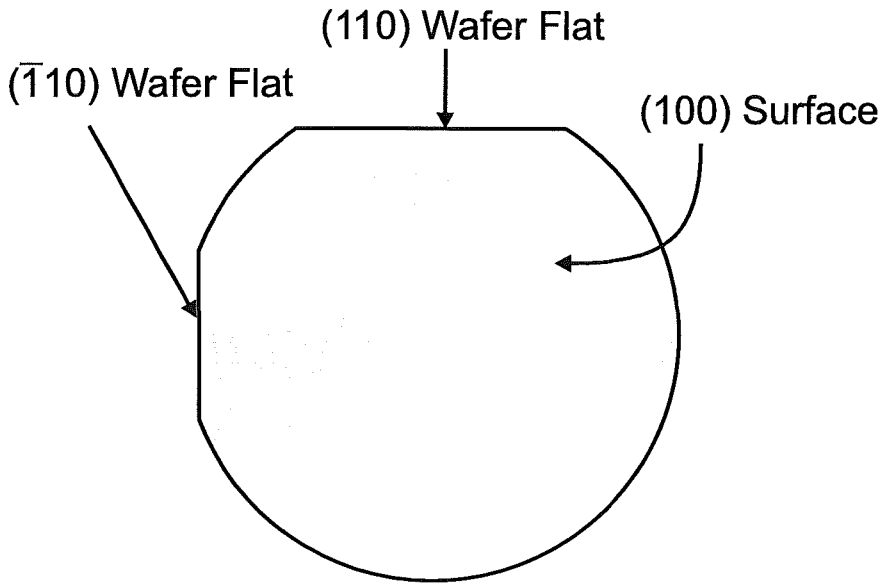


Figure 3.8 Wafer flats on the (100) silicon wafer.

V-Grooves are easily etched in (100) silicon, by using the {111} etch stop planes as the groove walls. To do this, the mask lines are oriented along the surface traces of the {111} planes, by aligning them to the ⟨110⟩ directions indicated by the wafer flats. In KOH solutions, the etch rate ratios of the low index planes {110}:{100}:{111} is reported to be 160:100:1 at 20°C [146]. The ratio increases to 400:200:1 at 85°C, and decreases to 50:30:1 at 100°C.

An isolated opening in the etch mask will eventually etch to a pyramidal pit bounded by {111} planes that exactly enclose the opening, as shown in Figure 3.9. A linear grating mask will produce V-grooves such as the one shown in Figure 3.10, as long as the mask lines are exactly aligned to the traces of the {111} planes on the {100} surface. The surface traces of the {111} planes intersect the (100) surface on the diagonal of the unit cell which corresponds to the ⟨110⟩ directions. Therefore, the mask can be aligned with the ⟨110⟩ wafer flats for an approximate alignment during whole wafer processing schemes. In some cases, a more precise alignment is required, or the wafer flats are lost when the wafer is cleaved into samples, in which case special alignment marks can be created by wet etching. This is described further in section 4.4. The aspect ratio of the V-grooves is fixed by the angle of the {111} sidewalls to the wafer surface, which can be calculated from simple geometry to be 54.74° for (100) wafers.

It should be noted that the use of multiple etch steps allows a wider range of geometries to be fabricated, such as cavities with vertical walls [147]. However, these are not described here because such process complexity is not required in the fabrication of the VADR plates.

3.7.5 Wet Etch of (110) Silicon

(110) silicon wafers are also widely used in micromachining. The surface is one of the {110} planes, and the wafer flats indicate the ⟨110⟩ directions. This wafer is useful because the {111} etch-stop planes are oriented vertically with respect to the wafer surface. However, their surface traces are no longer at right angles to each other, and result in parallelogram features with an obtuse angle of 109.47° as shown in Figure 3.11. In order to align a mask with the surface traces of the {111} planes, it must be rotated by 125.26° with respect to the wafer flat [140]. With these wafers it is possible to make rectangular grooves, as shown in Figure 3.12, which would be useful for the fabrication of the rectangular-toothed VADR devices described in Chapter 7. Note that both Figures 3.11 & 3.12 show a (110) plane at the bottom of the etched feature. This is accurate for short etches, but for longer etches the bottom of the features become aligned to the faster etching {100} planes, which are at an angle of 45° to the wafer surface. Note also that for the cavity etch shown in Figure 3.11, that longer etches or a large opening would result in the structure developing from a rhombus shape (shown)

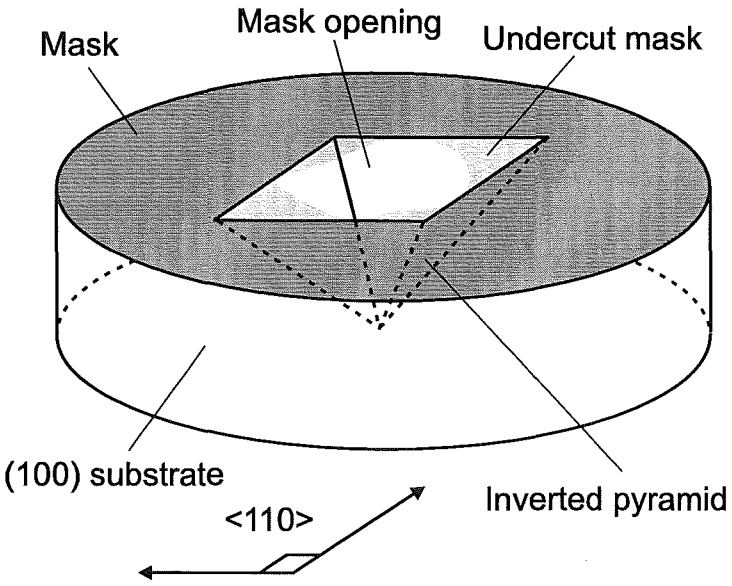


Figure 3.9 An arbitrary mask opening etches to an inverted pyramidal pit bounded on four sides by {111} planes.

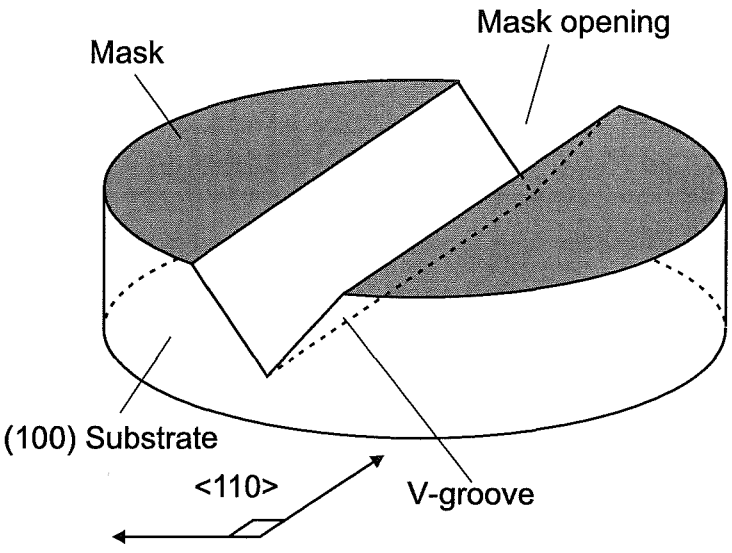


Figure 3.10 A linear mask opening aligned to [110] directions etches to a V-groove bounded on two sides by {111} planes.

into a hexahedron shape, with the two additional sidewalls forming in the acute-angled corners at an angle of 35.26° to the {110} surface plane [148].

3.7.6 Mask undercutting

Mask undercutting may occur during the anisotropic etching of both (100) and (110) silicon wafers. It has a significant impact on the final geometry of the etched features, and it must be well controlled in order to achieve good yields. The extent of the undercut is linearly dependent on the angle of the mask's misalignment (the angle between the edges of the mask lines and the surface traces of the {111} planes) [144,149]. The phenomena is illustrated in Figure 3.13, for the case of etching a V-groove in (100) silicon; note that the VADR plates are fabricated from such grooves. In Figure 3.13(a), the mask is perfectly aligned with the surface traces of the {111} sidewall (or etch-stop) planes, and it is not undercut at all. In Figure 3.13(b), however, the mask is grossly misaligned, leading to substantial undercut of almost half the final depth of the groove. The flares at the ends of the groove arise as the exposed convex corners at the edges of the wafer are etched. It is important to be able to predict how much undercut can be expected for a given misalignment, or vice versa, so that suitable alignment techniques and mask dimensions can be selected.

Undercut rates have been reported for masks that were up to seven degrees off alignment on (100) silicon [144], and up to three degrees on (110) silicon [149]. The data indicates that the undercut is slightly less severe for (100) silicon, although still significant. Over the measurement range, which is large enough to cover most actual situations, the ratio of the undercut U, to the groove depth d, can be expressed as a function of the angle of misalignment θ

$$\frac{U_{(100)}}{d} = \frac{\theta}{40}, \tag{3.7a}$$

$$\frac{U_{(110)}}{d} = \frac{\theta}{35}. \tag{3.7b}$$

The implication of undercut rate is that, as the required etch depth increases, better alignment and / or wider mask lines are required. For example, assuming that a mask could be aligned to within 2° of the wafer flats on (100) silicon and the wafer flats were themselves only accurate to 1°, there would be a total alignment tolerance of ±3°. In order to etch V-grooves to a depth of 200μm, the mask lines would have to be at least 30μm wide. If it was desired to reduce the mask lines to 5μm thick, the alignment accuracy would have to be improved to better than ±1°. Accurate alignment and careful choice of mask dimensions, using equation 3.7a or 3.7a as a guide, is required to obtain good yields, particularly for deep etches.

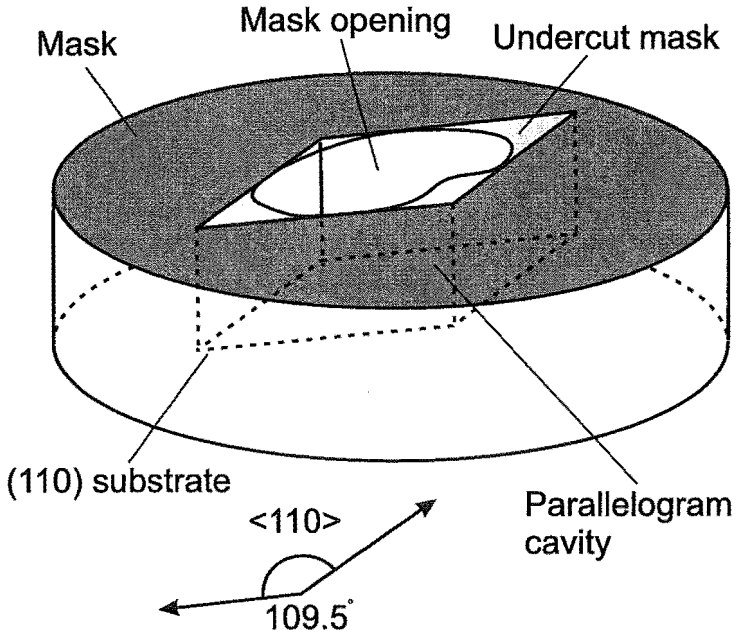


Figure 3.11 A small arbitrary mask opening on (110) silicon etches to a parallelogram (rhombus) cavity bounded by vertically oriented {111} planes on the sides. The bottom of the cavity is not bounded by an etch stop plane. Note that for larger openings and longer etch times, a hexahedron is formed instead of a rhombus.

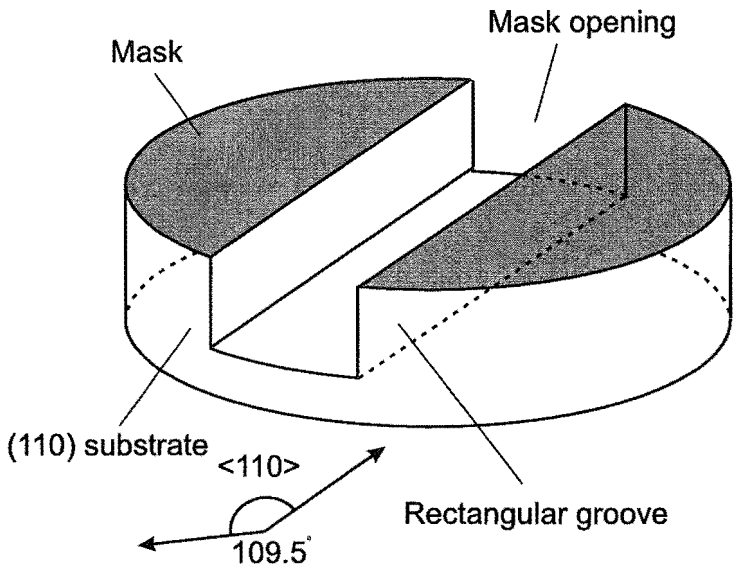
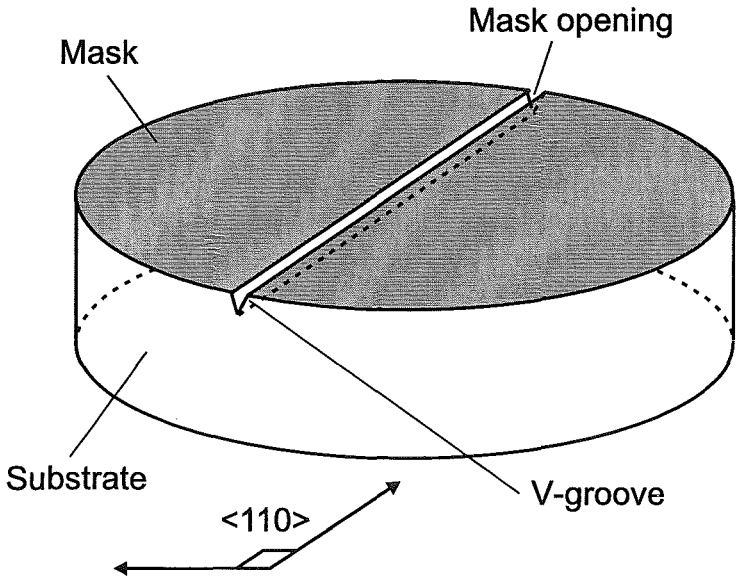
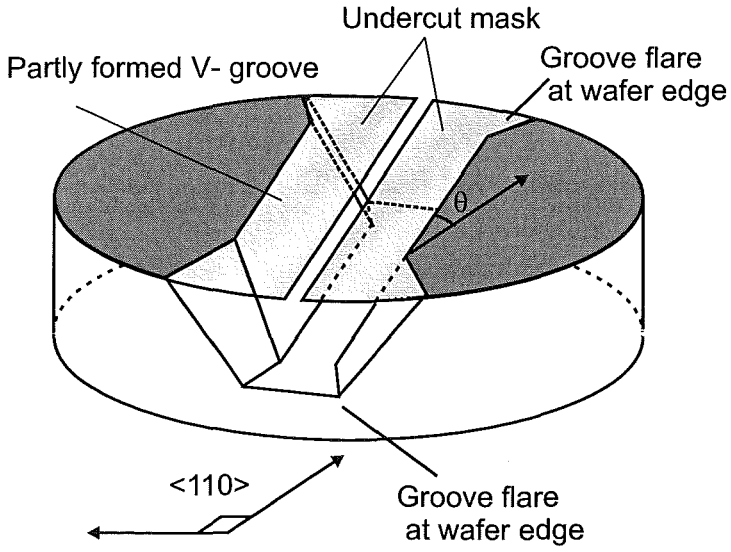


Figure 3.12 A linear mask opening on (110) silicon, aligned to [110] directions, etches to a rectangular groove bounded on two sides by {111} planes.



(a)



(b)

Figure 3.13 Mask undercutting when etching (100) silicon. If the mask is perfectly aligned to the ⟨110⟩ directions then there is no undercutting (a), whereas there is significant undercutting if it is misaligned (b).

3.7.7 Wet Etch Equipment

The wet chemical etching apparatus is shown in Figure 3.14. A custom-blown glass flask was used to contain the etchant, and it is pictured in Figure 3.15. While KOH attacks glass, the reaction is sufficiently slow that the flask has survived three years of use already. The main cylinder of the flask is 95mm in diameter and 80mm in height. There were three openings in the top - one each for the reflux column, thermometer and sample holder. The main section of the reflux column was 330mm in length, with 25 turns of 10mm tubing for the water cooling. Tap water was slowly flowed through the reflux, from the bottom to the top. The reflux was essential for maintaining the etchant concentration during long etches. A retort stand and clamp were used to steady the reflux. The thermometer was used to check that the etchant temperatures were in agreement with the readout of the waterbath temperature control unit. The sample holder was a PTFE stick, which was held in the centre of the flask by a red rubber bung placed in the central opening. The stick had several slits in the end, allowing the silicon samples to be mounted horizontally in the flask.

The flask was immersed in a Grant Instruments Y-14 controlled-temperature water bath that maintained the temperature at 80±1°C through the use of an electrical heating element. Grant PS-20 polypropylene balls, 20mm in diameter, were floated on the water bath surface in a single layer to reduce evaporation during long etches. Evaporation of the water, and subsequent fall in water level in the water bath risked a reduction in the temperature of the etchant in the flask. The water depth was maintained at approximately 115mm throughout the etch. The etchant was stirred by a 25mm Teflon-coated magnetic stirring bar in the flask, which was actuated by a water-powered magnetic stirrer motor placed under the flask. The stirrer was powered by a water pump that was built into the waterbath. Etching was conducted in a fume cupboard, although the reflux ensured that little, if any, KOH escaped into the atmosphere.

3.7.8 Hillock formation and elimination

The exact mechanism of the formation of hillocks, such as those shown in Figure 3.16, is presently the subject of debate [150]. While the debate is of great interest elsewhere, because it is closely related to the determining the underlying mechanisms of anisotropic etching in general, this thesis is concerned with a more immediate practicality - preventing hillock formation. Nonetheless, a brief description of the current state of thought is warranted. Following this is a description of an etchant formulation for the prevention of hillocks that was determined from the literature and tested by the author.

As shown in Figure 3.16, hillocks are pyramidal structures, with a definite crystallography. Their size increases with age, and over the course of a long etch, it is

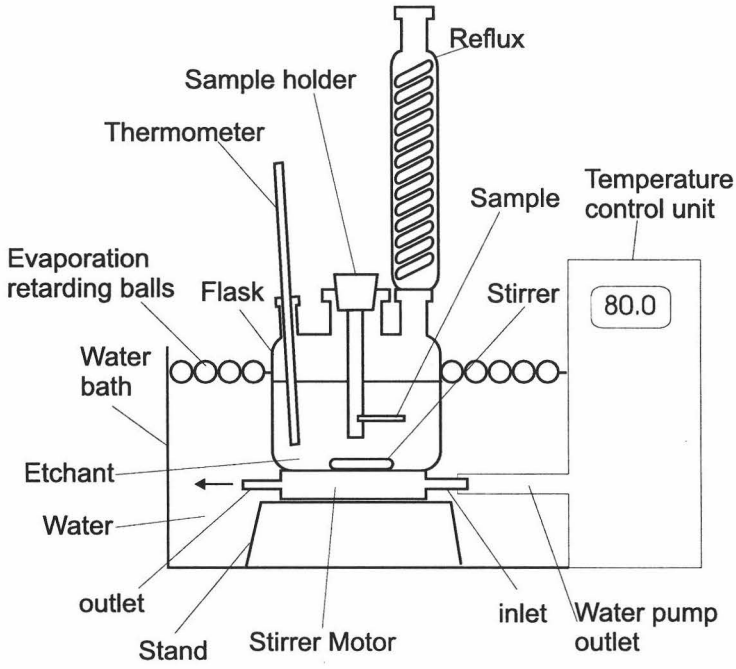


Figure 3.14 Refluxed wet anisotropic etch flask in waterbath

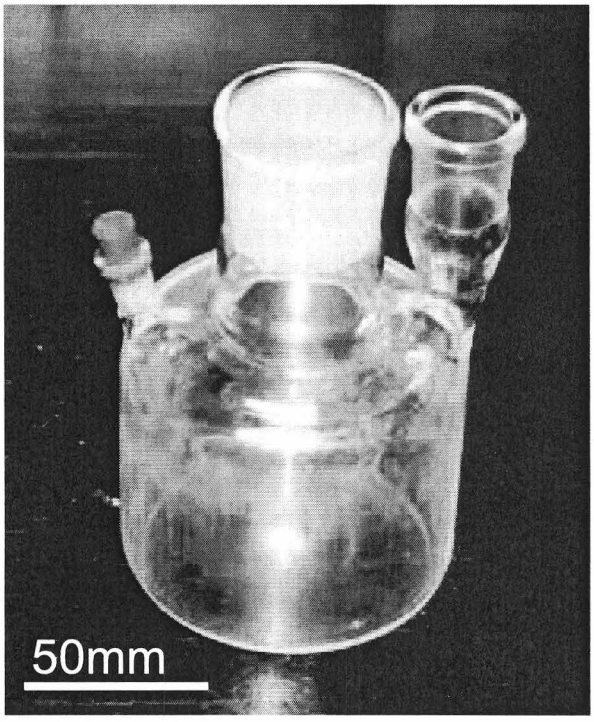


Figure 3.15 The custom-blown glass flask used for the alkaline wet etching of silicon, in the UoC laboratory.

not unexpected to find a range of sizes. The greatest mystery of the hillock is that it ought not exist. If it is assumed that the hillocks are composed of perfect surfaces, each etching at a rate determining solely by the crystal orientation of that plane, then it becomes impossible to explain their presence because they ought to etch from the top down at, at least, the (100)-etch rate. This is obviously not the case, and it is generally accepted that, through some mechanism, the top of the hillock locally retards the etch rate. The nature of mechanism that is the subject of the debate. Several authors have speculated that the deposition of etching products or SiO_2 precipitates on top of the hillock is responsible. Others speculate that the hydrogen bubbles evolved during etching are responsible for creating micromasks where they stick to the surface, and that the silicon etch reaction is running in the reverse direction, resulting in regrowth.

A recent paper closely observed hillocks and concluded that they were not caused by defects in the bulk crystal. They then presented computer simulations that supported their hypothesis that semipermeable particles (such as silicate colloid particles) sticking to the top surface, and ridges, of the hillocks are responsible for their formations [150]. The debate continues.

One earlier paper, [151], showed experimentally the effects of temperature, KOH concentration, and solution agitation on hillock density. Samples etched in solutions that were agitated had approximately half the hillocks of those in solutions that were not. A series of experiments with 15% wt/wt KOH showed that hillock density logarithmically increased with increasing temperature, although no temperature had zero hillock density. Finally, it was shown that increasing the concentration of KOH, reduced the hillock density. The temperature was maintained at 80°C, and once the concentration was increased to approximately 40% wt/wt, the hillock density was zero. No information was given on the etch rate of this solution, except to say that they accepted trends for etch rate variation with temperature and etchant concentration could not explain the changes in hillock density. Since some variation in etchant preparation is usually encountered between laboratories, due to slight differences in the water content of the solid KOH and the quality of the water, a nominal 40% wt/wt solution was prepared for testing. 133.4g KOH pellets were added to 200ml of DI H_2O, resulting in a moderately exothermic reaction. The solution was heated to 80°C, and silicon samples were etched with a variety of arbitrary features. Etches were repeatedly hillock free, so there was no need to adjust the etchant formula. The (100) etch rate was measured to be 1.25μm/min, which compares favourably with other reported etch rates, as shown in Table 3.2.

3.8 SILICON ETCH SIMULATION

Computer simulations of etch processes provide a useful tool for testing novel mask designs before committing them to expensive photolithographic masks. In this thesis,

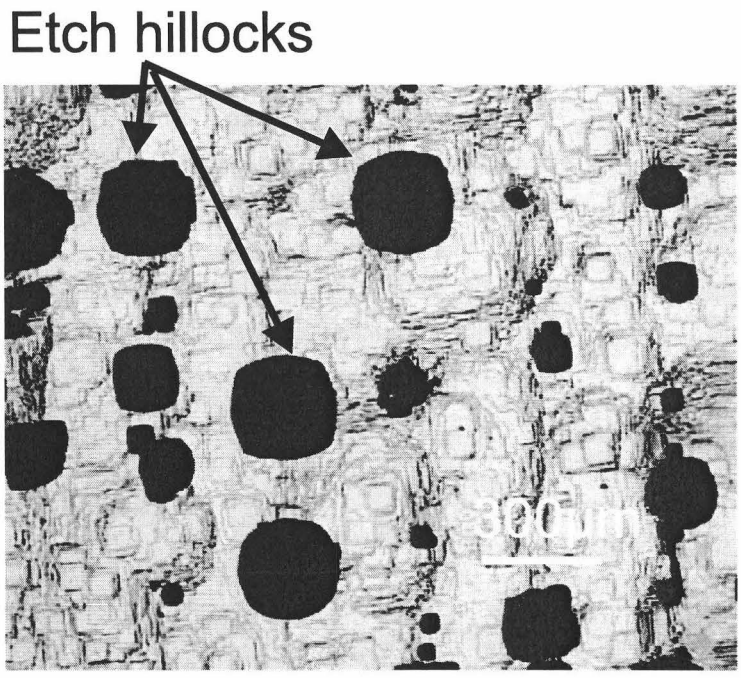


Figure 3.16 Example of hillock formation during wet etching. The hillocks vary in size according to their age.

Table 3.2 Selected reported KOH etch rates

Rate μm/min	Concentration % wt/wt	Temperature °C	Author
0.90	15	72	Seidel *et al.* [146]
1.25	40	80	This thesis
1.40	33	80	Williams *et al.* [141]

the Anisotropic Crystalline Etch Simulator (ACES) [152] was used to evaluate patterns designed to reveal the orientation of crystallographic planes in silicon. There are two approaches to modelling crystalline etching - the geometric method and the cellular automata method. The geometric method treats the substrate's surface as a continuous entity, while the CA method represents it as a large number of individual cells organised in a suitable lattice. Since ACES uses the continuous cellular automata (CA) method, that is what will be described in this section [153].

Since the substrate is silicon, the CA cells are organised in a diamond lattice. Each cell contains information on the strength of its bonds to the neighbouring atoms, which determines when it is etched (removed from the lattice). Unlike reality, the CA cells may assume partially etched states. This allows ACES to model the arbitrary etch rates required to accurately represent actual anisotropic etch processes, without artificially roughening the etch surfaces like non-continuous CA methods.The use of many cells

leads the CA method to use more memory than the geometric method, however it is highly efficient and accurate when modelling elaborate mask shapes.

Rather than model the chemistry of an etch solution, which is currently unknown in many cases anyway, empirically determined values of the relative etch rates of the {111}, {110}, {100} and {311} planes are specified. This allows great flexibility in modelling different etch processes, whether they be isotropic dry etches or anisotropic wet etches. ACES's default KOH solution models a solution similar to that of Williams *et al.*in Table 3.2. For modelling etches in this thesis, the relative rates were modified to better represent the measured (100) etch rate of the 40% wt/wt solution used at 80° (see section 3.7.8). Note that the rates for the other planes were not measured for this etchant, but were estimated using an anisotropy of 1:100:200:200 for the {111}:{100}:{110}:{311} planes, that is well within the maximums reported (see section 3.7.3). The relative etch rates of both the default and the modified ACES 'solutions' are shown in Table 3.3.

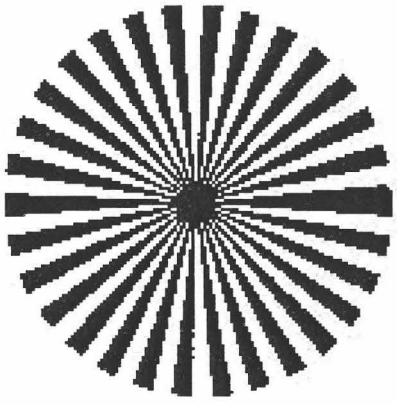
In order to show the effects of mask orientation, and reveal the relative etch rates of the crystallographic planes, a spoke pattern may be etched. A coarse version of this spoke, 148μm by 146μm in size, is included in the ACES package. It is shown in Figure 3.17(a), with the results of etching a (100) silicon wafer in the default KOH solution for 10min, with the spokes as openings in (b) and with the spokes as masks in (c). The angle-dependent underetching of the masks is clear in both cases, and it is the ability of ACES to predict this behaviour of masks at arbitrary orientations to the crystal lattice that is of use in the etch modelling conducted in this thesis.

Table 3.3 Relative etch rates for KOH solutions in ACES's

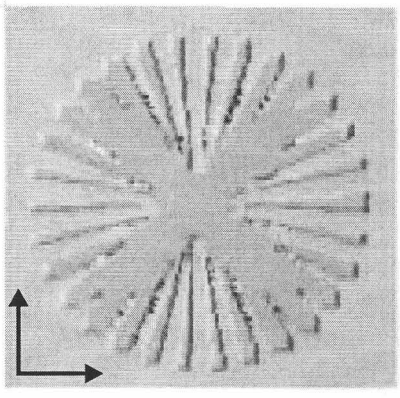
Solution	{111}	{100}	{110}	{311}
default	0	1.0	1.4	1.4
40% wt/wt, 80°C	.0125	1.25	2.5	2.5

3.9 SUMMARY

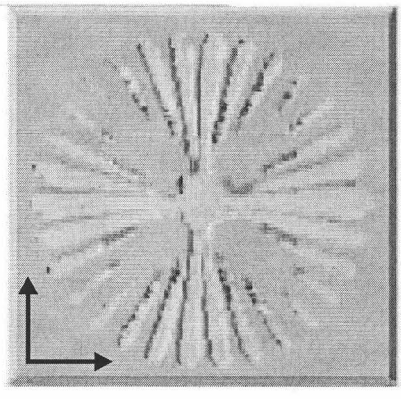
This section has described the principles and equipment, of the processes employed in the fabrication and inspection of VADR plates. Photolithography was used to define patterns in photoresist, and a number of related activities were described, including the process of making chrome-on-glass masks via the photoplotting of reticles, and the process of spinning, exposing and developing both positive and negative photoresists. Details of IRL's Mann GCA PG3000 photoplotter and UoC's Cobilt CA800 mask aligner were also provided. The UoC's Oxford Plasmalab 80+ RIE machine was described along with the process of etching Si_3N_4 in a plasma of CHF_3 and Ar gases. Anisotropic wet etching of silicon, with potassium hydroxide, is the most important



(a)



(b)



(c)

Figure 3.17 ACES etch simulation. The spoked mask in (a) was used to etch (100) silicon for 10 min, with the spokes as openings in (a) and with the spokes as masks in (b). The arrows were applied afterwards, and indicate the ⟨111⟩ directions.

part of the fabrication process because it allows the otherwise-difficult geometry of the V-grooves to be produced accurately and routinely, with a high quality finish. The lattice structure of silicon and the method of identifying crystal planes using Miller indices were reviewed. The particular geometries that can be produced in a wet etch depend on the orientation of the crystal lattice to the wafer surface, and the etching of (100) and (110) wafers was described, along with a summary of the current state of thought regarding the anisotropic wet etch mechanisms. A crystalline etch simulation technique was introduced, that utilised predetermined relative etch rates for four major planes, to model wet etching. This aided in the development of the alignment marks. The wet etching equipment, including a custom-blown glass etch flask, was also described. Hillock formation and prevention was briefly reviewed.

Chapter 4

VADR FABRICATION

4.1 INTRODUCTION

This chapter describes the details of the fabrication of VADR devices for use at 100 GHz, using the microfabrication techniques introduced in the last chapter. While small V-grooves may be readily wet-etched in silicon, fabrication of the large grooves (500 μm period) required by the VADR plate design presented additional difficulties. In particular, precise alignment of the mask to the crystal planes was critical, to avoid excessive undercut and destruction of the grooves. Unfortunately, due to equipment limitations, it was not possible to employ the standard procedure of aligning the V-groove photolithographic mask to the wafer flats, followed by processing of the whole wafer at once. As a result, much of the effort was absorbed in developing a procedure for individually determining the orientation of the crystal planes on each of the samples cleaved from the original silicon wafer.

4.2 VADR PLATE DESIGN

The VADR device for 100GHz is a 10× scaled-up version of the device proposed for use at 1 THz (see section 2.5.1) [65]. The device consists of two identical V-grooved plates; an example of one is depicted in Figure 4.1. The plates are up to 25 mm by 25 mm in size (due to process limitations), and are patterned on the top surface with parallel V-grooves of 500 μm period and 350 μm depth, as shown in the cross section of Figure 4.2. The two plates are intended to interlock as shown in Figure 4.3. Note that the key aspect of the VADR device is that the separation distance, s, between the two interlocked plates can be varied.

In order to ensure the 100 GHz device performed like the proposed 1THz device, the scaling maintained the feature-size to wavelength ratio of the grooves. The groove period was $\frac{\lambda_0}{\sqrt{3}\times n_{Si}}$ and the depth $\frac{\lambda_0}{\sqrt{6}\times n_{Si}}$, where λ_0 was the free space wavelength at the design frequency and n_{Si} was the magnitude of the refractive index of the silicon substrate. Although there was a slight difference between the originally-assumed refractive index of silicon at 1THz, $n_{Si}(1\text{THz}) = 3.46$ [65], and the measured value at

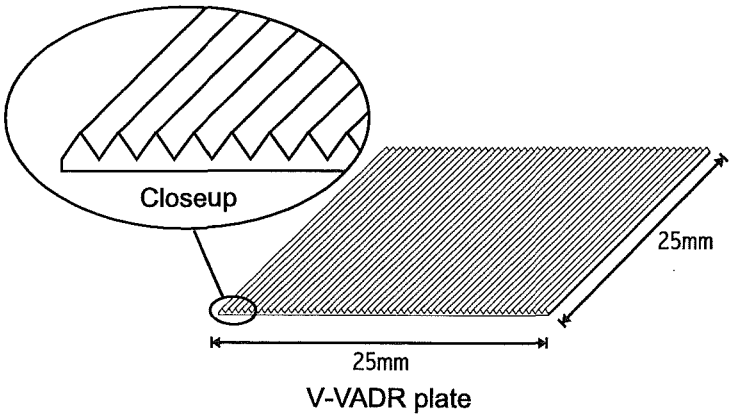


Figure 4.1 The desired VADR plate is up to 25mm by 25mm square and the top surface is patterned with parallel V-grooves.

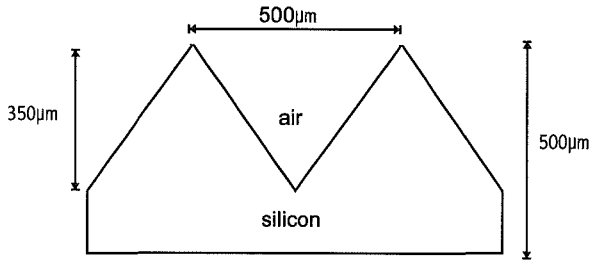


Figure 4.2 The plate is 500μm thick, the grooves are 350μm deep and have a period of 500μm.

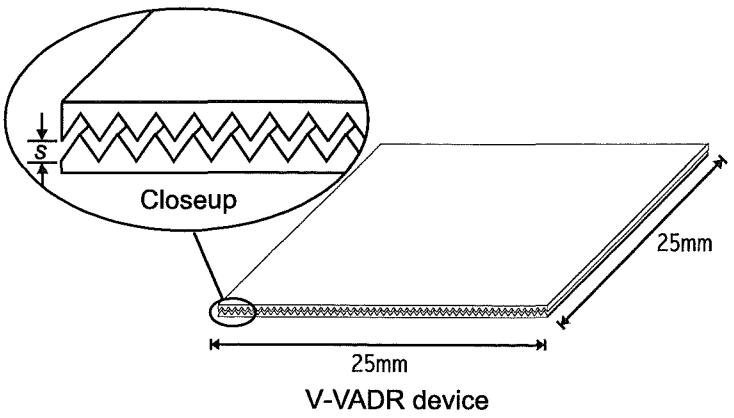


Figure 4.3 Two interlocked VADR plates form a VADR device. They are shown here in a partially interlocked position, where s is the separation distance.

100GHz, $n_{Si}(100\ \text{GHz}) = 3.42$ [16], this was not factored into the calculation of the new feature size since it was not expected to appreciably affect the device's performance and it would allow convenient, round numbered, groove dimensions to be retained. The groove dimensions for the device at both frequencies are summarised in Table 4.1.

Table 4.1 VADR groove dimensions at design frequencies of 1THz and 100GHz.

Frequency	n_{Si}	Wavelength μm	Groove depth μm	Groove Period μm
1 THz	3.46	300	35	50
100 GHz	3.42	3000	350	500

The ratio of the desired groove period to depth was $\frac{1}{\sqrt{2}}$, which was ideally suited to fabrication in (100) Silicon, where {111}-bounded V-grooves having the same dimensional ratio are easily produced. The large size of the grooves did present some issues, particularly with regards to selecting the width of the mask lines for the final etch step that produces the grooves. Due to the undercutting phenomena described in section 3.7.6, wide mask lines are required if the alignment accuracy is poor. Otherwise, the mask lines cannot withstand the severe undercutting associated with inaccurate alignment. If they become completely undercut, the groove tips are exposed to the etch solution, and the grooves are ultimately destroyed.

However, overly wide mask lines would hinder the device's performance. If the tips of the grooves were too flat, and the bottoms of the grooves had been etched to a fine point, then there would be large air gaps between the plates at full interlock, preventing the device from being able to achieve zero retardance, or in other words, preventing it from being able to be "turned off". If etching was stopped prematurely, in order to give matching flat bottomed trenches, the situation at interlock would be somewhat improved due to the reduced air gap, however the maximum retardance of the device would suffer slightly from having shallower grooves. Additionally, since the alignment accuracy is poor, there could be substantial variations in the amount of misalignment from sample to sample, which could make it difficult to create a pair of plates with similarly-sized flats on the tips and bottoms of the grooves.

For this proof of concept device, a perfect performance is not necessary, and the maximum acceptable tip size, and hence mask line width, is set at a somewhat arbitrary 40μm, equivalent to 8% of the period. A mask for making the VADR plates, with lines of this thickness, was prepared. The mask pattern is reproduced in Figure 4.4. It is worth noting that, although plates with larger tips would make an inelegant VADR device, it would still operate so long as the two plates were able to interlock. In order to routinely fabricate the VADR plates with the 40μm mask lines, it was necessary to improve the precision of the alignment over that obtained by aligning to the rough side of a cleaved sample. The required improvement is quantified in section 4.4, where it is

described how the wet etching process can be used to both reveal the directions of the {111} planes, and in the same action, produce mask alignment marks.

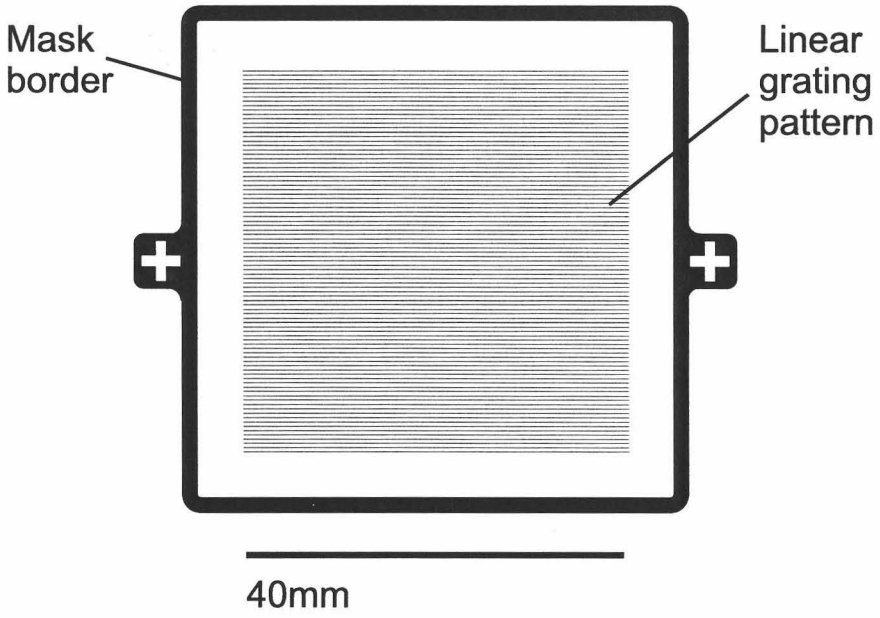


Figure 4.4 The VADR mask consisted of a linear grating with 40μm thick lines, on a period of 500μm. The mask surround was a standard feature of IRL's process.

The consequence of using an alignment mark was that the usable sample size was reduced, because the area where the marks were patterned was not able to be used to make grooves. The maximum plate size was limited in this fabrication process, by the neck of the wet etch flask, to 25mm by 25mm. Since the birefringence of the device was under examination, it was important that the device was square so that the same usable plate area was presented to the radiation beam of the test equipment with the grooves oriented either vertically or horizontally. Since the 3dB beam width was expected to be 20mm at the location of the device under test (see section 5.3.2), it was important that the final VADR plate be not less than 20mm by 20mm, and certainly larger if possible. Thus, the maximum area that could be used for the alignment mark was 5mm by 25mm.

4.3 FABRICATION PROCESS

The VADR plate fabrication process can be broken into two major stages. First, the crystal orientation is established by briefly wet etching an alignment mark to reveal the ⟨110⟩ directions. Second, the VADR mask is aligned with the aid of the etched alignment mark, and the V-grooves are formed with a long wet etch. In the following process outline, both of the major stages are broken into smaller steps, and presented in

the order in which they were performed. Note that the two stages are quite similar, and that details of the process mechanisms, equipment and process recipes can be found in the relevant sections of Chapter 3.

Stage 1 – Establishment of the ⟨110⟩ crystal directions

The steps of stage 1 are illustrated in Figure 4.5.

Step 1: Scribe and cleave wafer

Coating the Si_3N_4 layer on the front (polished) side of the wafer protects it from contamination during scribing. Typically, 1ml of Shipley S1813 photoresist is spun on in the photoresist spinner, followed by a short bake of 15 minutes at 90°C. Longer bakes and higher temperatures make the resist more difficult to remove later. The wafer scriber described in section 3.3 is used to mark the 100mm diameter wafer into 25mm by 22mm samples, which are then cleaved. The sample is made rectangular, so that when the 3mm thick alignment mark is later sacrificed, the VADR plate becomes square (22m by 22mm).

Step 2: Define the alignment mark pattern

The alignment mark is a shape that has been optimised to clearly show the direction of the important silicon crystal planes when anisotropically wet etched. In the final process, the dual radial alignment mark is patterned in AZ5206 image reversal (negative) photoresist on the front side of the wafer. The geometry of the alignment marks is further explained in section 4.4.

Step 3: Dry etch the alignment mark pattern into the Si_3N_4

RIE is used to etch the exposed Si_3N_4, with the photoresist acting as an etch mask. The etch parameters are detailed in section 3.5.2.

Step 4: Remove passivating substances from the etch surface

An oxygen plasma etch for 10-20 minutes is performed to remove any passivating (etch-inhibiting) organic substances from the exposed silicon surfaces.

Step 5: Short wet etch of the alignment mark

The alignment mark is etched in a 40% by weight solution of KOH in de-ionised (DI) H_2O, with the Si_3N_4 acting as an etch mask. The etchant is maintained at a constant temperature of 80°C throughout the 80 minute etch. This etch is anisotropic, and intended to reveal the orientation of the crystal planes by the variation in etch rate that is a function of opening geometry and orientation. Simulated and actual etch results are presented for the two types of alignment mark trialled by the author, in sections 4.4.2 and 4.4.3.

Step 6: Cooling and cleaning

After the etch, the sample is transferred to a beaker of DI H_2O to cool, to prevent the formation of native oxide on the exposed silicon surface. This is a precaution in

case further etching of the alignment mark is required. The sample is rinsed in H_2O to remove any traces of the KOH etchant, and dried with flowing N_2.

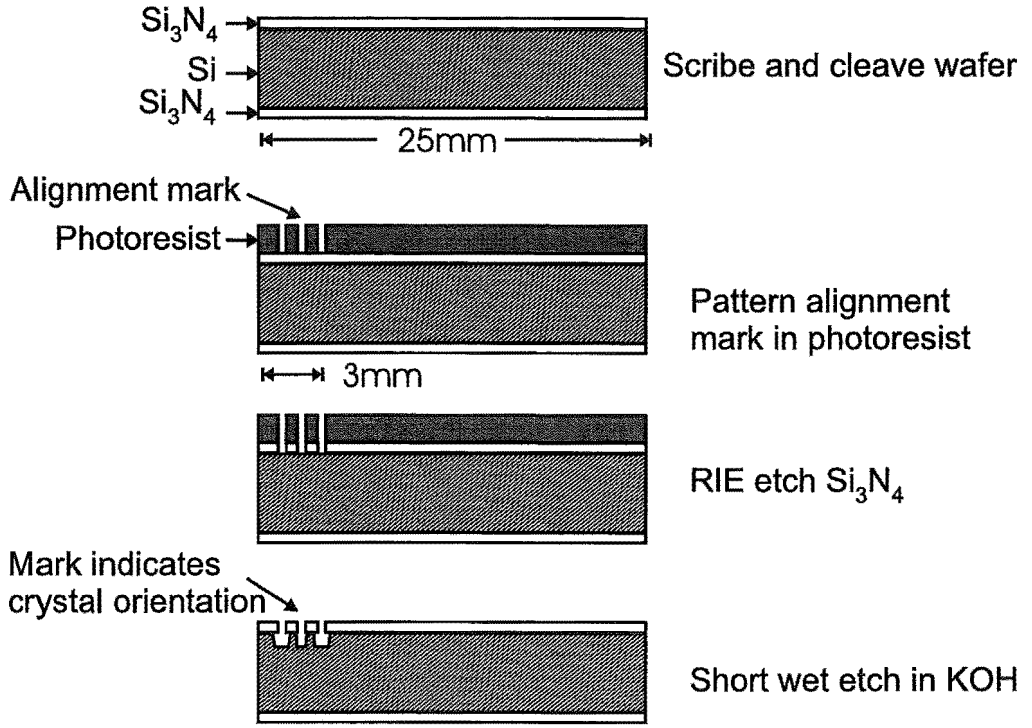


Figure 4.5 The first stage of fabrication involves cleaving the samples and independently establishing the {111} crystal plane orientation with a short pre-etch.

Stage Two – Formation of the V-grooves

The steps in stage two are illustrated in Figure 4.6.

Step 1: Define the linear grating

The linear grating mask of Figure 4.4 is patterned on the wafer surface with S1813 photoresist. The linear grating is aligned to the ⟨110⟩ directions, as established by the alignment marks of section 4.4.3.

Step 2: Dry etch the linear grating into the Si_3N_4

The linear grating is patterned into the Si_3N_4 using the etch recipe in section 3.5.2, i.e. the same etch process as in Stage 1, Step 3.

Step 3: Remove passivating substances from the etch surface

An oxygen plasma etch for 10-20 minutes is performed to remove organic substances from the exposed silicon surfaces which may inhibit the etch.

Step 4: Wet etch of the linear grating

The V-grooves are formed by the preferential etching of the {100} and {110} planes over the {111} planes. The sidewalls of the grooves are formed by the {111} planes. The etch results are presented in section 4.5, with a discussion concerning the elimination of pyramidal hillocks, in section 3.7.8.

Step 5: Remove remaining Si_3N_4 mask and backside protection

The remaining silicon nitride coating on the front and backside is removed by immersion in 40% by weight HF acid, at room temperature, for 20min. The VADR plate is now complete.

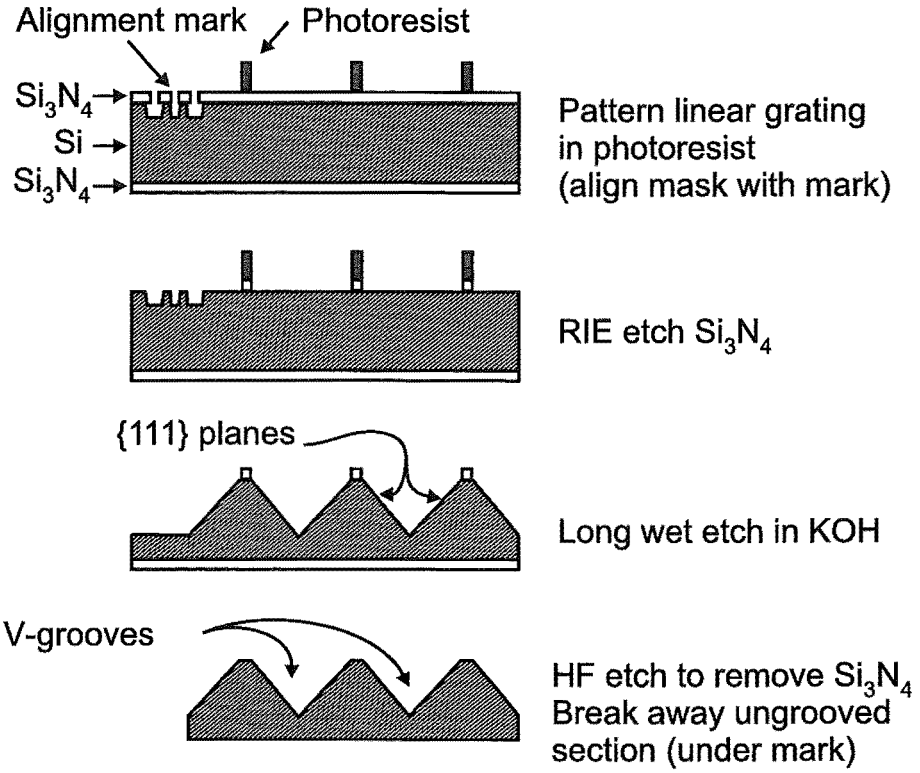


Figure 4.6 In the second stage of fabrication, the V-grooves are etched into the top surface of the VADR plates.

4.4 ESTABLISHING THE CRYSTAL ORIENTATION

This section describes how a short wet etch can be used to establish the crystal orientation on each of the samples, in order to improve the mask alignment accuracy and ultimately make sharper V-groove tips. The required alignment accuracy is calculated from the desired mask line width, and groove depth, in order to aid in the selection of suitable pattern for the alignment mark. Existing mask patterns are reviewed, and

evaluated for use in this fabrication process on the basis of pattern simplicity, required precision in the pre-alignment, overall size and accuracy of the crystal orientation indication. Two variations on these patterns were trialled by the author, using both computer simulation and actual etch experiments. From the results of these, the mask line thickness of the VADR plates was chosen.

The mask alignment accuracy on the samples, without the aid of an alignment mark, is $\pm 4°$, comprising the $\pm 1°$ error in the wafer flats, the $\pm 1°$ error in scribing the sample sides parallel to the wafer flats, and the $\pm 2°$ error in aligning the mask to the usually rough sample sides. According to Equation 3.7a, for an etch to a depth of 350μm, and a mask $4°$ off alignment, there would be undercut of approximately 35μm on each side of the mask line. Thus the mask line would need to be wider than 70μm in order to avoid becoming completely undercut, perhaps $75 - 80$ μm. This mask line is greater than the 40μm mask line width decided on in section 4.2. If the alignment accuracy were able to be improved by a factor of two, to $\pm 2°$, then only 35μm of undercut would have to be withstood, and the mask with the 40μm wide lines would be able to be used.

The alignment mark would need to be able to cover a range of at least $8°$, to cope with the $\pm 4°$ accuracy in aligning to the sample sides. It would also need to occupy a strip of one side of the wafer of no more than 5mm by 25mm, as discussed in section 4.2. With these requirements quantified, the suitability of existing patterns can be evaluated.

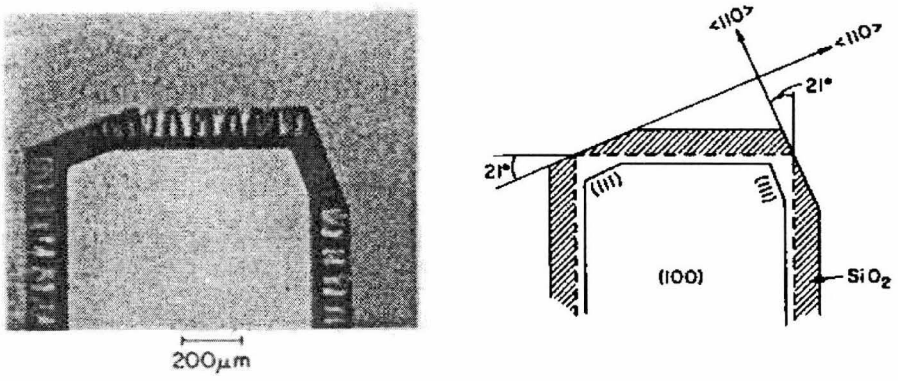
4.4.1 Mask review

The need for greater precision in mask alignment than can be afforded by the wafer flats has prompted several authors to develop mask patterns, that when etched briefly, will clearly and accurately reveal the $\langle 110 \rangle$ directions of the crystal lattice within the wafer [146,148,154–157]. These patterns are also of use when the direction of the wafer flat is not exactly known, as is the case on the cleaved samples used in this thesis to fabricate VADR plates. Four representative patterns are described in this section, in the order they were reported in the literature. Each of them draws on a different aspect of the wet etch process to produce their indication of the $\langle 110 \rangle$ directions.

The first pattern comes from E. Bassous' 1978 paper, where he suggested the use of a greatly misaligned rectangular mask opening, such as the one shown, after etching, in Figure 4.7(a). The opening of Figure 4.7(a) has been misaligned by $21°$, and the formation of (111) planes in the corners is evident. The relevant planes are identified in the drawing of Figure 4.7(b). The size of the (111) planes increases as the etch progresses, until eventually the feature is bounded solely by (111) planes. This pattern has the advantage of not needing to be closely aligned to the expected $\langle 110 \rangle$ directions. In fact, the opposite is true - a generous misalignment in the order of $10 - 30°$ makes

the (111) planes in the corners more distinct. A disadvantage is the short distance over which the masks may be aligned, the length of a {111} corner plane.

The second pattern has been reported by several authors, and can also be employed to characterise the relative etch rates of an etchant [146, 154, 155]. It is a fan-shaped pattern, with mask openings that appear to represent the spokes of a wagon wheel, such as the one shown in Figure 3.17. This pattern relies on the fact that the lateral etch rate is strongly dependent on the alignment of the mask, with an etch rate minimum at the ⟨110⟩ directions. The mask is simply aligned with the spokes that are least undercut. In practice, many more spokes are used to improve the precision of the indication, which may be as good as ±0.1° for a large pattern. Such a pattern, if it was to use 5μm thick spokes, with no taper, and allowing for a 15μm arc-length space between the tips of the spokes, would have a diameter of 23mm, almost as big as the VADR plate.



(a) (b)

Figure 4.7 A misaligned rectangular opening develops {111} planes in the corners, to which a mask can be aligned, photograph of etched opening (a), schematic of etch planes (b). (Reproduced from [148] Figure 3.)

The third pattern is almost a combination of the previous two. G. Ensell suggested in his 1995 paper the use of 75μm diameter circular openings [157], placed along an arc of 45mm radius, as shown in Figure 4.8(a). This pattern was etched until the circular openings had etched fully to inverted pyramids (see section 3.7.4). Then, the ⟨110⟩ directions were determined as shown in Figure 4.8(b), by identifying the neighbouring inverted pyramids that had the least translation misalignment. The circular openings are drawn in the figure with an angular spacing of 1°, but the actual pattern had an angular spacing of 0.1°. Two arcs were used, with a common arc-centre, so that subsequent masks could, with the aid of a split field microscope, be more accurately aligned than in the case of a single alignment pattern. This pattern is large, requiring a 90mm by 3mm area while only allowing for half the required pre-alignment accuracy. In order to allow for the ±4° pre-alignment precision, the pattern would need to be nearer

6mm across. At the reported arc radius, δx increases by 0.14μm for each $0.1°$ increase in $\delta\theta$, however, if the arc radius was reduced to 10mm, so that the pattern would fit on a VADR plate, δx would increase only at the rate of 0.030μm per $0.1°$ increase in $\delta\theta$. It is possible that at the reduced pattern size, the indicated $\langle 110 \rangle$ directions may prove difficult to discern.

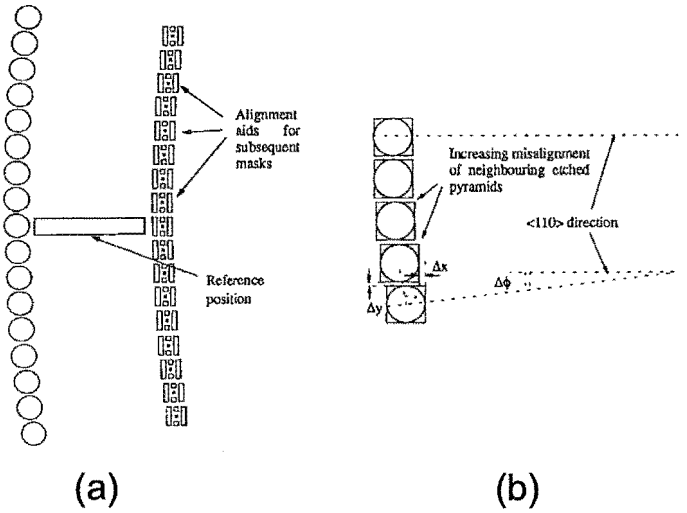


Figure 4.8 An array of circular openings located on a 45mm radius arc (a), etch to form inverted pyramids shown schematically in (b), that show increasing translational misalignment between adjacent features as the angular misalignment to the $\langle 110 \rangle$ directions increases. (Reproduced from [157] Figures 2,3.)

The last pattern to be covered in this review provides the best precision. In the 1996 paper of Vangbo *et al.*, they claim that the eye is more sensitive to symmetries, and suggest the use of an array of fork-shaped patterns with triangular prongs, as shown in Figure 4.9. The prongs of the fork pattern etch to different lengths when misaligned, and only match each other in length when the fork is accurately aligned. The arrow in Figure 4.9 indicates the fork that corresponds to the $\langle 110 \rangle$ directions. The marks were angularly spaced by $0.1°$. Each of the forks is 600μm long, and 150μm thick, so this pattern would require approximately twice the area of Ensell's pattern, 12mm, in order to cover the range of the pre-alignment precision required by the VADR plates.

It is clear that the reduced area available on the VADR plates would prevent the use of the fan, Ensell's or Vangbo's patterns without settling for poorer precision or making modifications. The only pattern that appears immediately suited is that of Bassous, and an implementation of this is trialled in section 4.4.2. Due to the drawback already identified with this pattern, a further pattern is trialled, that is a compact form of the fan pattern. This dual radial alignment mark is described in section 4.4.3.

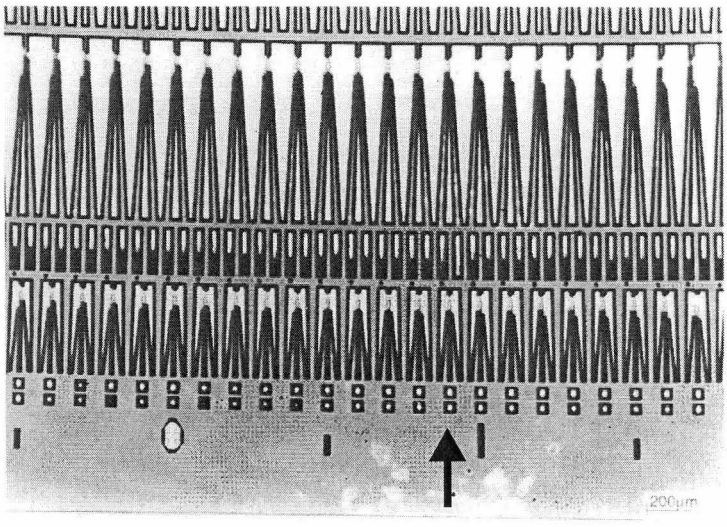


Figure 4.9 Mask alignment mark using an array of fork patterns. The symmetrically etched fork pattern, which is aligned to the [110] direction, is indicated with an arrow. (Reproduced from [156] Figure 4, arrow added.)

4.4.2 Square Window

The square window mask, shown in Figure 4.10, provided for a neat implementation of Bassous' misaligned-rectangle alignment pattern, as there were many corners in which the {111} planes could form. Conveniently, the mask was found in an existing set of test patterns, and it could be trialled without incurring extra mask making costs. The whole feature was 2mm in outside dimension, and constructed from 200μm lines. The four square openings in the window were 700μm in dimension.

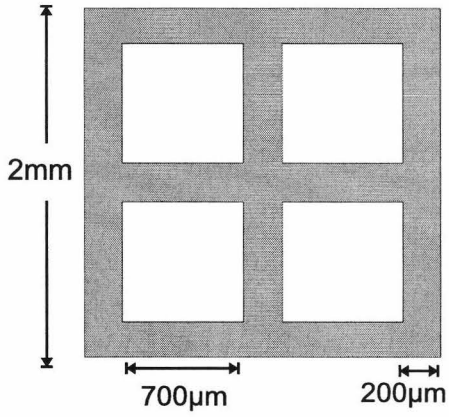


Figure 4.10 Schematic of the square window mask pattern

Before etching, the ACES computer program (see section 3.8) was used to simulate

the results. The pattern was simulated at one tenth the size, to speed the simulation. Since each pixel represents an area 1μm $\times 1\mu$m, a bitmap mask was prepared that was 247 by 247 pixels, with the side of the window feature being equivalent to 200 pixels long. The pattern was misaligned by 16°. After 8min of simulated etching using the 40% wt/wt KOH at 80°C etch model, the geometry of the sample was plotted, as shown in Figure 4.11. The original mask edges have been overlaid to show how the {111} planes are tangential to the enclosed corners of the mask. Four of the sixteen {111} planes have been indicated by arrows, although their presence is clear. Their length is 32μm in the simulation. It is expected then, that after etching the actual mark for 80min, the {111} planes would be 320μm long.

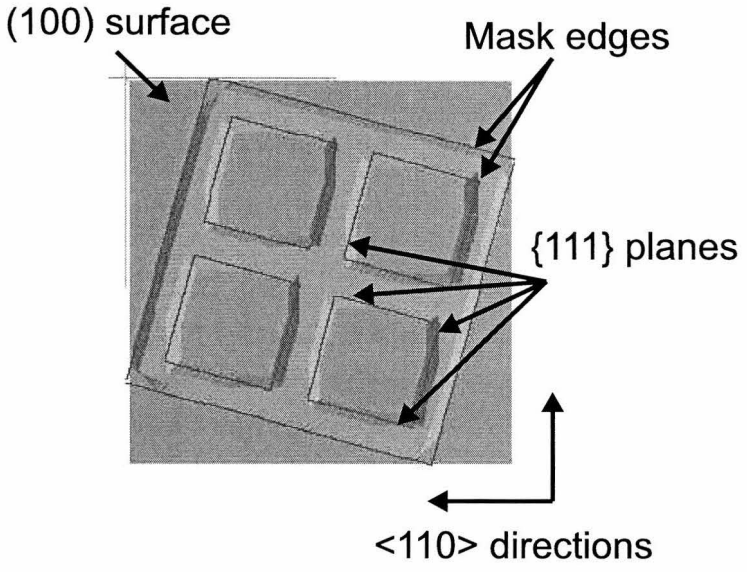


Figure 4.11 ACES etch simulation of the square window alignment mark

An actual sample was prepared and etched using the procedural steps of stage 1 in section 4.3. The mask was misaligned by 16° as it was in the simulation. The result of the etch is presented in 4.12. The {111} etch stop planes are exposed at the corners, and are 300μm long, which is within 7% of the length predicted by the simulation. With the 100× optics on the Cobilt mask aligner, it was possible to align the marks to $\pm 15\mu$m over the length of one of the corner planes (300μm), giving rotational alignment to better than ±3° of the {111} planes. This is an improvement of 1° over aligning to the edges of the samples. The first successful plates were etched with the aid of this alignment mark, but the ability to routinely fabricate plates with this alignment mark was hindered by the modest precision.

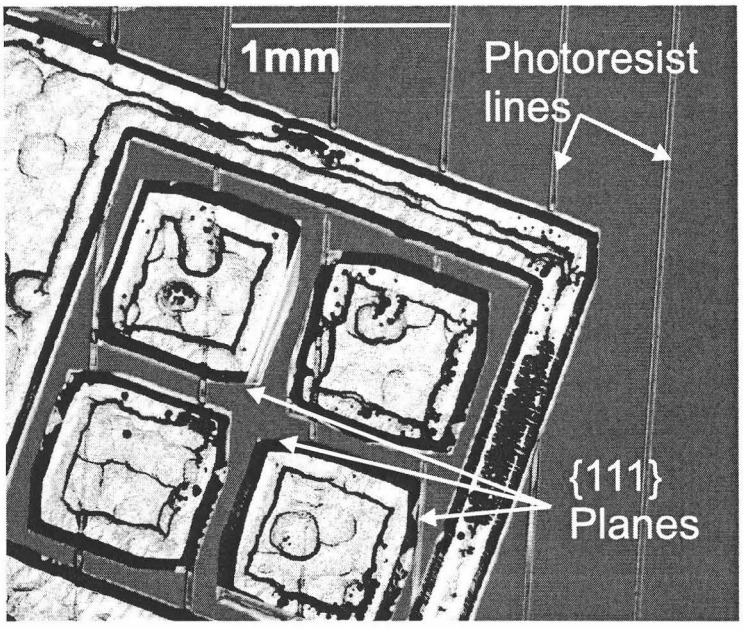


Figure 4.12 Photograph of the etched square window, with linear grating patterned in photoresist over the top.

4.4.3 Dual Radial Mark

The dual radial mark, shown in Figure 4.13, was based on a description of a pattern that had been used at IRL, but for which the mask and pattern data were not available [158]. It exploits the strong variation in lateral etch rate as a function of angular alignment. The mask underetch rate is at a minimum when the opening is aligned to the $\langle 110 \rangle$ directions, so the subsequent masks need only be aligned to the least undercut spoke on the pattern. Since the two sets of spokes share a common centre, it is possible to align each end of a single mask line to the spokes with the aid of the mask aligner's split field microscope, providing greater precision than was possible with the single pattern of the square window. At a misalignment of 4°, the lateral etch rate is in the vicinity of 8μm/hr. Two variations of the mark were produced on the mask, each with 41 segments per side, on an arc of radius 23mm, but with a different angular spacing. Mark one had an angular spacing of 0.2°, allowing for a pre-alignment accuracy of 8°, while mark two had 0.4° spacing to allow for 16° pre-alignment accuracy. Each spoke was 500μm long, and 10μm thick, except for every fifth segment which was 700μm long.

Before etching, the ACES computer program (see section 3.8) was again used to simulate the results. The pattern was simulated at a tenth the size, to speed the simulation. Since each pixel in the bitmap mask represents an area 1μm $\times 1\mu$m, a bitmap mask was prepared that was 105 by 140 pixels, with the length of the large spokes being equivalent to 70 pixels long. The spokes were angularly spaced by 0.4°,

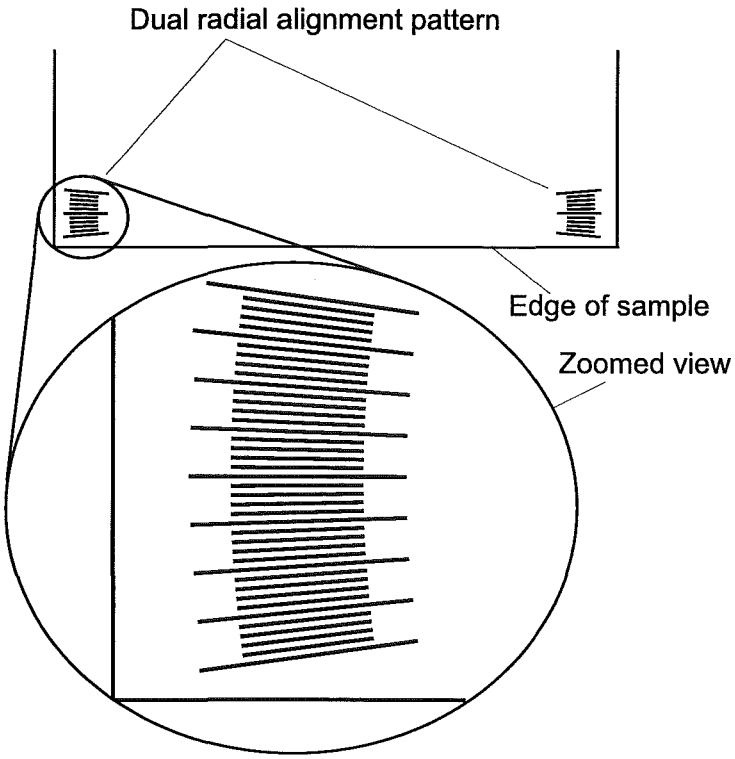


Figure 4.13 Dual radial alignment mark as patterned on a VADR plate, with zoomed view of the left hand half of the pattern.

the same as for mark two. After 8min of simulated etching using the 40% wt/wt KOH at 80°C etch model, the geometry of the sample was plotted, as shown in Figure 4.14. The original mask edges have not been overlaid, to better show the variation in the width of the etched spokes. It is clear that the central spoke is the least underetched, which is to be expected since the ⟨110⟩ directions were aligned with the sides of the bitmap mask. It is expected then, that after etching the actual mark for 80min, there would be an appropriate amount of variation in the width of the etched spokes. It is important to avoid overetching, or otherwise all of the etched spokes may merge, obscuring the true alignment of the crystal lattice.

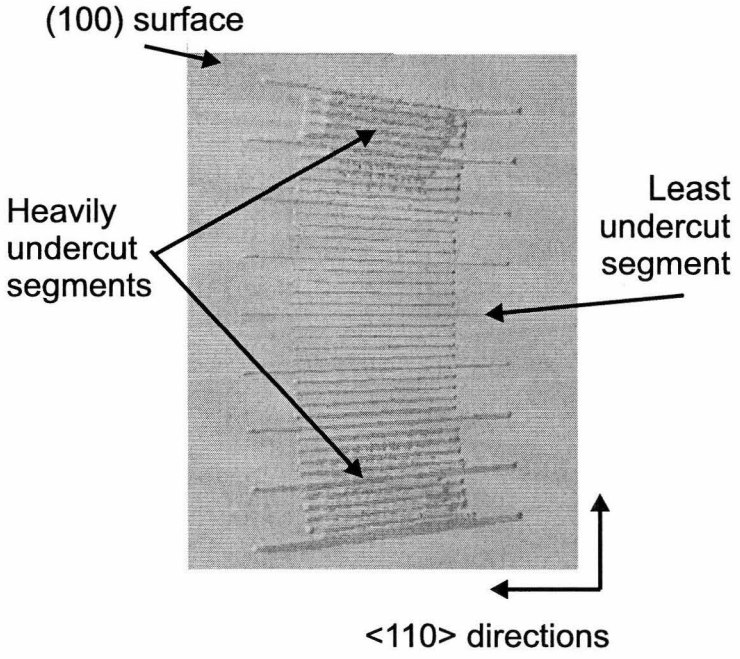


Figure 4.14 ACES etch simulation of one half of the radial alignment mark.

A test set of each mark was etched for 80 min in 40% w/w KOH at 80°, giving the results presented in Figure 4.15(a,b). As in the simulation, the [110] direction is clearly indicated by the least undercut segment. With the 100× optics on the Cobilt mask aligner, it was possible to align the mask lines to within ±15μm of the marks, over a distance of 20mm, giving the precision of alignment to the marks of ±0.4°, to give an overall precision of better than ±1° in ideal circumstances. This pattern was also compact, requiring only a 3mm wide section of the sample to be sacrificed, less than the 5mm allowed. The consequence of restricting the pattern to such a small size is that the precision suffers. If this was a problem in further experiments, the spoke spacing of the radial alignment mark could be reduced. As it was, the dual radial alignment marks were used to routinely fabricate the VADR grooves, so no further improvement

in alignment precision was attempted.

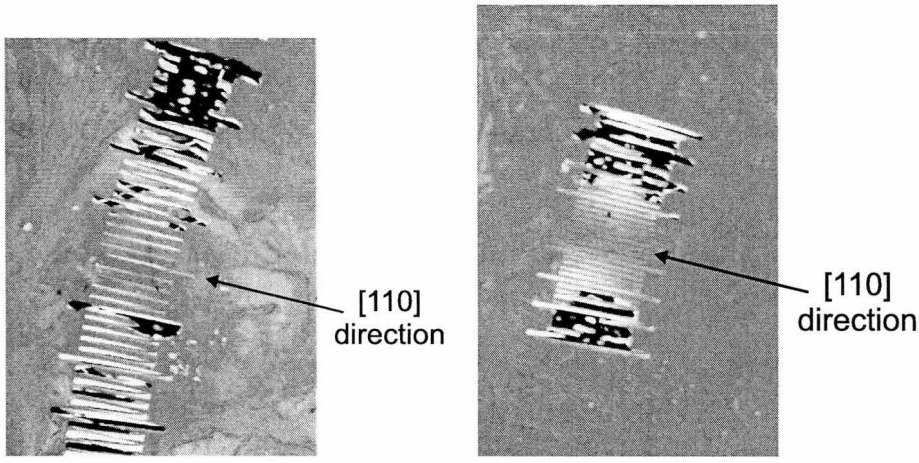


Figure 4.15 Dual radial alignment marks after 80min etching in 40% w/w KOH at 80°, 0.4° angular spacing (a), 0.2° angular spacing (b).

4.5 RESULTS OF PLATE FABRICATION

This section presents scanning electron microscope (SEM) images of the completed VADR plates, that have fabricated with the dual radial alignment marks and the 40μm line width VADR mask, and discusses how various aspects of the fabrication process have affected the outcome. In total, nine plates were successfully fabricated using this process, with no failures.

A completed plate is shown in the SEM micrograph of Figure 4.16. The tips are 10μm, and the bottoms 20μm, in width. Since the sidewall angle is known, the groove depth can be calculated to be 332μm. Thus the actual (100) etch rate for this etch was 1.19μm/min, which was within 5% of the earlier estimated etch rate of 1.25μm/min. The difference in etch rate is attributed to variations in the preparation of the etchant. The lateral underetch rate was 0.054μm/min. The ratio of undercut to depth is 0.045, and using Equation 3.7a it is calculated that the mask was misaligned by 1.8°. This was greater than expected, and it was attributed to an alignment error such as aligning one mask line to the correct spoke on the left side, while unintentionally aligning an adjacent mask line to the correct spoke on the right. This would contribute a 1.4° error, and if the spokes were, by chance, exactly aligned to ⟨110⟩ directions, then the only error would be 0.4° in the alignment to the spokes, totalling to give the observed 1.8° misalignment.

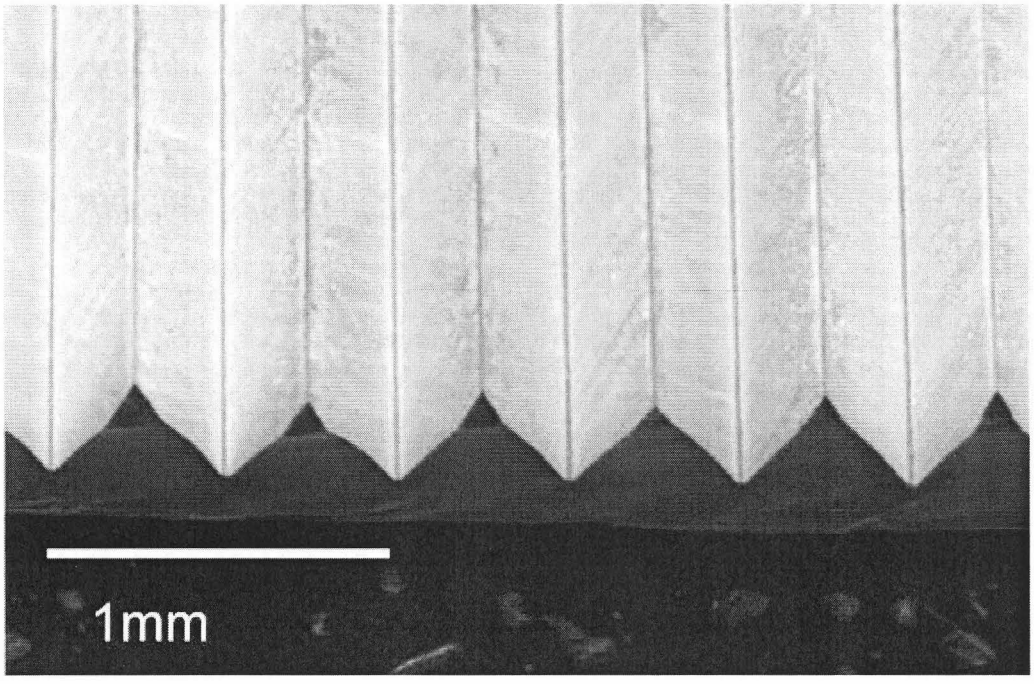


Figure 4.16 SEM micrograph of completed VADR plate.

A striation or ripple effect is visible on the sidewalls of the grooves. This is attributed to the formation of misorientation ledges [149], a consequence of the mask misalignment. The nanometre-scale ledges do not affect the ability of the plates to interlock, and in some cases may help to prevent the plates from sticking together. To the eye, the surfaces appear as a mirror finish. Another effect, not pictured, was a 2μm ledge evident every 3mm on every groove. This was a result of a stitching error at the reticle photoplotting stage. The possibility of adjusting the machine to eliminate this was discussed with IRL, who owned the machine, however their advice was that it was a difficult operation, likely to do more harm than good. As it was, the 2μm stitching error was insignificant, did not prevent interlock, and was thus acceptable at this stage of the device's development.

Two plates are shown fully interlocked in Figure 4.17. This is one of the positional settings of the VADR device, and it is important that the fit is as close as possible so that the device may be 'turned off'. In this case, there was a gap of 20μm between the plates. This level of fit is acceptable for the proof of concept device, and quite reasonable given that the plates were patterned and etched separately.

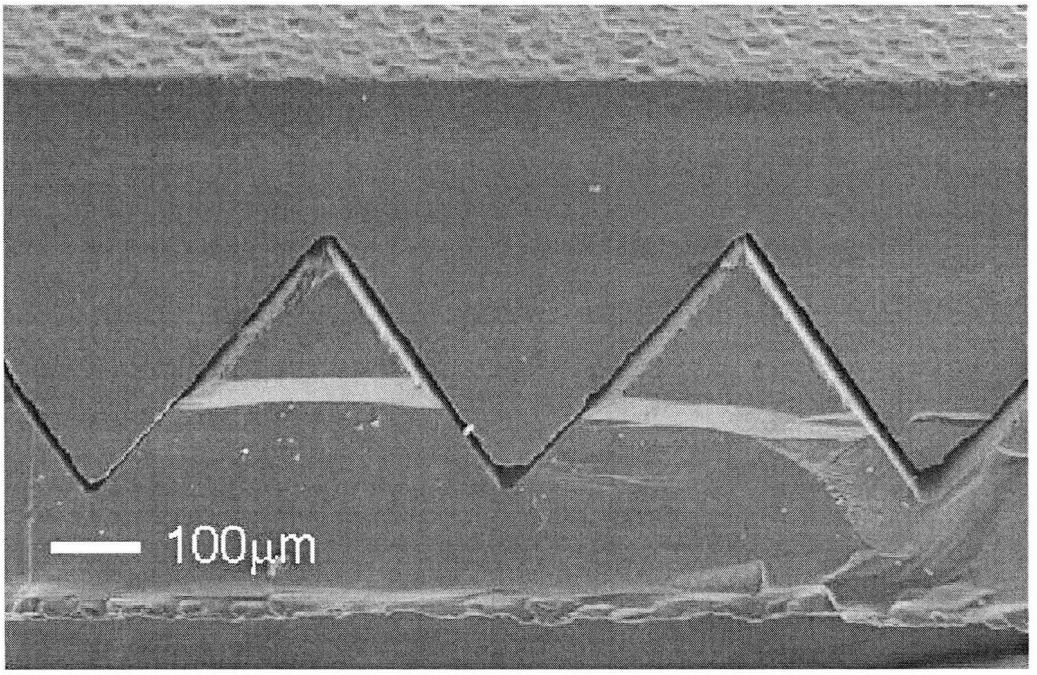


Figure 4.17 SEM micrograph of two fully interlocked VADR plates. A 20μm gap is evident. Note that some grooves appear to have formed perfectly, while some appear to have flattened tips.

In order to improve the quality of the plates, there are several steps that could be taken. Specifically, the improvements could be made to the repeatability of the groove profile, so that a better fit could be achieved between two plates that are fully

interlocked (i.e. no separation between the two plates), and to the sharpness of the groove tips and bottoms, so that finer control of retardance at small separation distances is possible and the maximum retardance is increased.

Currently, the sample size is constrained by the neck of the etching flask to less than 25mm by 25mm. This means that on every sample, the crystal orientation must be separately established, and the photolithography separately performed. Then, the grooves must be separately etched. Being able to perform any or all of these activities simultaneously on more than one sample would improve repeatability, and hence the plates would mate better at full interlock . One approach to improving repeatability would be to pattern the entire wafer with the linear grating, using the wafer flats for alignment. The entire wafer could then be dry etched to make the Si_3N_4 wet etch mask, and then diced into the into 25mm by 22mm plates, each of which would then be separately wet etched. This would give a mask alignment precision approximately equal to that actually achieved in the existing plates (1.8°), without the need to sacrifice any of the plate area to alignment marks. This approach was not adopted earlier in order to conserve materials.

A potential problem with this approach lies in the photoplotting stage. While the photoplotter will accept commands to move over an area of 111mm by 111mm, the available reticles were limited to 76mm by 76mm, and within that, the patternable area was further reduced to around 60mm due to the presence of IRL's mask surround which was automatically added to any pattern. The presence of this mask surround could be avoided by editing a converted PG file by hand, or by directly creating the PG file from scratch. It is assumed that by not using the mask surround, enough mask area could be patterned to justify the use of a whole wafer at once.

In order to improve the groove profile, three approaches may be taken. Either the plates may be overetched, a narrower mask line can be used, or a deliberate misalignment can be introduced to cause a controlled amount of mask undercut. Without the development of suitable in-situ inspection methods, it would be difficult to achieve a good yield from an overetch process. Both of the remaining methods require much more precise mask alignment, and this is only possible if an alignment mark such as the one suggested by Vangbo is used. To obtain the $\pm 0.05°$ precision reported for a KOH solution, it would be necessary to use the full-size pattern, and this in turn would require that the wafer remained intact throughout the process. The mask making equipment, photolithography tools and the RIE machine are all capable of handling 100mm wafers. The only equipment not capable is the wet etch flask. Fortunately, the water bath used to heat the flask is large enough to accommodate a larger etch flask of suitable design, and the capability exists at the University of Canterbury (in the Chemistry department) to create such a flask, from glass. Once the alignment marks were etched and the linear grating was patterned, or even after the V-grooves were etched, the wafer could be diced into smaller plates. The narrower-mask-line approach will

yield the smoothest sidewalls, as the misorientation ledges will be either non-existent, or much smaller. This may or may not be desirable, depending on how much stiction the smooth surfaces would introduce when the plates were fully interlocked.

Since the plates already produced were suitable for testing at the proof of concept stage, the improvements mentioned here have not been implemented. Should it become necessary to produce further plates, a great improvement in the groove profile and the fit of the plates could be realised with a small financial outlay to cover a new etch flask and photolithography masks.

4.6 SUMMARY

This chapter has described the procedure used by the author for fabricating VADR plates. Where possible, standard UoC laboratory procedures were used, as described in the previous chapter. Some modifications were required to optimise the fabrication process, but once the final process was developed, VADR plates were routinely fabricated. Several methods of establishing the orientation of the crystallographic planes were reviewed, and trials of square window and dual radial alignment marks were conducted by computer simulation and experiment. The dual radial alignment mark was the better of the two, and gave alignment to within $\pm 1^\circ$, including the precision of the crystal direction indication and the precision of aligning to the mark. The 40μm mask line width of the VADR mask's linear grating (used to pattern the V-grooves) resulted in flat tips at the top of the grooves, although this was perfectly acceptable for the proof of concept device and likely only to affect the performance at small separations. Two interlocked plates were imaged in order to measure the gap, which was 20μm was evident. Using the fabrication process described in this chapter, nine plates in total were produced, with no failures.

Chapter 5

VADR MEASUREMENTS

5.1 INTRODUCTION

This chapter presents measurements of the silicon VADR device at W-band frequencies, but particularly at the design frequency of 100 GHz. These free-space measurements are an essential step in the validation of the VADR concept. Even with the best of equipment, it is a particularly challenging task, and care needs to be taken to ensure accurate results.

In this chapter, aspects of the experimental design, equipment and procedures are discussed, and measurement results are presented. The measurements were conducted with state-of-the-art W-band free-space measurement equipment made available at the University of Glasgow, augmented by special fixtures that were developed to mount and actuate the VADR plates in accordance with the experimental design.

5.2 EXPERIMENTAL DESIGN

The key experimental results required are a quantification of the retardance and insertion loss of the VADR device, since these quantities allow the operation of the device to be predicted for any input polarisation (in conjunction with the Jone's calculus, or Stokes parameters described in Appendix A.2). A number of approaches may be taken in determining the behaviour of a device that operates on the polarisation state, and three possible methods are described here, in order to make clear the advantages of the method that was actually used (the complex transmission coefficient method).

5.2.1 Direct polarisation measurement

The first method involves the direct measurement of the polarisation state input to, and output from, the DUT. This can be achieved by measuring the polarisation state at a receiver (RX) with and without the DUT present. The difference between the two polarisation states is equal to the polarisation change caused by the DUT. In order to measure the polarisation state, the detector must be a linearly polarised antenna

that can rotate. Alternatively, a rotating linear polariser such as a wire grid may be used in front of a non-linearly polarised antenna. For either detector arrangement, the unwanted cross-polarised field components must be strongly rejected (attenuated to less than 0.1% power, or -30dB). The source (TX) may take any polarisation state. A diagram of the direct method is shown in Fig. 5.1. In order to detect the polarisation state, the amplitude of the received signal is measured and recorded as the detector is rotated through 360°.

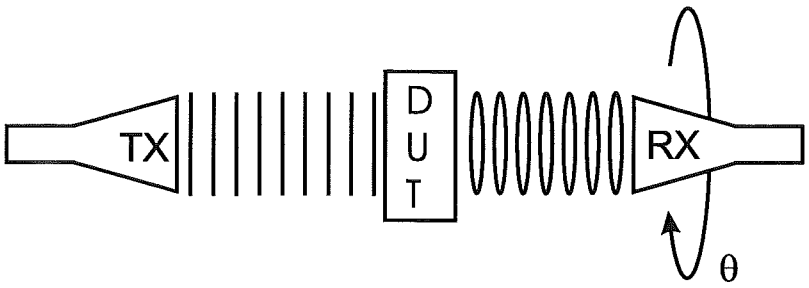


Figure 5.1 Block diagram of the setup for the direct method of determining the polarisation change caused by the device under test (DUT). A transmitting antenna (TX) produces a plane wave of known polarisation. The receiving antenna (RX) is rotated, and the electric field amplitude is measured for all angles.

The amplitude of the electric field is plotted against the angle it was measured, as shown in Fig. 5.2(a) for the case of an elliptically polarised beam. The magnitude shows two maxima and two minima, corresponding to the major and minor axes of the polarisation ellipse, respectively. The two maxima and the two minima are evenly spaced at angular increments of π radians, while each minima is spaced $\pi/2$ radians from either maxima. The two angles at which the maxima are produced give the angle of the major axis of the ellipse, and the amplitude gives the length of the major axis. The same applies to the minima for the minor axis. The re-constructed polarisation state is shown in Fig. 5.2(b). For linearly polarised beams, there are two maxima, and two nulls (depending on extinction ratio of detector), while for circularly polarised beams, the amplitude is same at all rotational angles. This method has the disadvantage of requiring a rotating receiver, which may be impractical in some cases where the receiver unit is bulky or sensitive to physical disturbance.

5.2.2 Comparison with known device

The second method indirectly measures the polarisation change caused by the DUT. In this method, the transmitter (TX) and receiver (RX) should be identically polarised; for this description, they are taken to be linear. The most important difference in the receiver requirements is that the receiver does not need to rotate. Instead, a calibrated polarisation compensator is included between the DUT and the receiving antenna (RX),

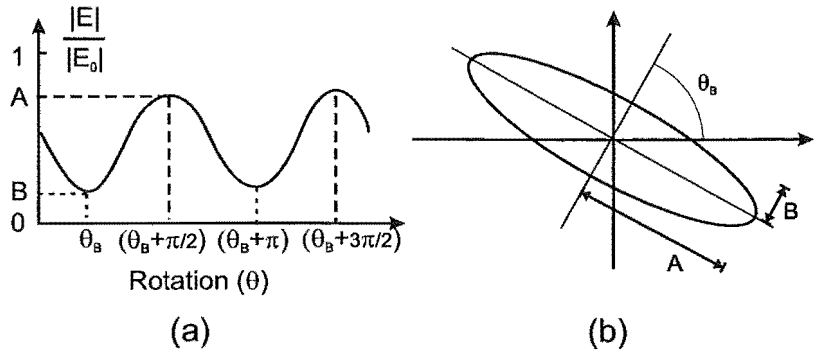


Figure 5.2 Sample results for the direct method of determining the polarisation state, in the case of an elliptically polarised beam. (a) normalised electric field amplitude as a function of receiving antenna angle, (b)reconstructed polarisation state.

as shown in the block diagram of Fig. 5.3. The transmitter and receiver are aligned the same, so that any variation in the polarisation of the transmitted signal (caused by the DUT) will cause a drop in the received power. The known compensator is initially set so that it does not alter the polarisation. The DUT is then interposed between the transmitter and known compensator, and set to a position of interest. The received power should drop as a result of the change in the polarisation state. To determine how much the polarisation has been changed, the calibrated polarisation compensator is adjusted until the amplitude of the received radiation again reaches its maximum (minus the insertion loss of the DUT). The polarisation change in the DUT is equal, but opposite to, the polarisation change from the known compensator. The point of using a known compensator is that the polarisation change can be easily determined by looking up its current positional or electrical control setting on its calibration chart. This technique is common at optical frequencies where there are wide variety of variable polarisation compensators, such as the Babinet-Soleil, and others described in section 2.3. However, it is lack of devices with this functionality at millimetre and sub-millimetre wavelengths that lead to the proposal for the VADR device in the first place. Eventually VADR (but more likely one of the improved versions developed in chapter 7) could be used as a known, calibrated, polarisation compensator.

5.2.3 Complex transmission coefficient measurement

The third method, the one adopted for the present experiment, is analogous to the simulation methodology employed in the VADR device proposal [65], where the complex transmission coefficient (i.e. magnitude and phase) is determined separately for transverse electric (TE) and transverse magnetic (TM) incident waves (see section 2.2.3 for a definition of the TE and TM waves). This method has the advantage of allowing a direct comparison between the measured and simulated transmission coefficients, fur-

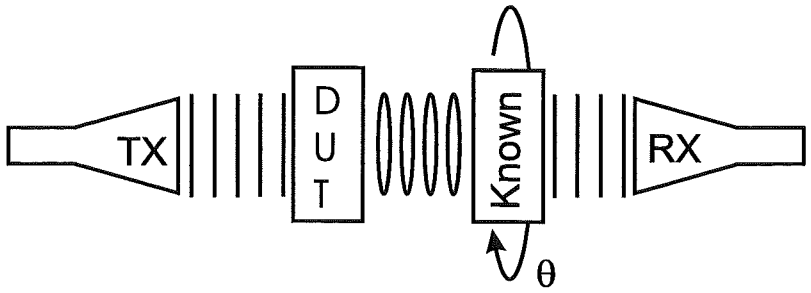


Figure 5.3 Block diagram of the setup for an indirect method of determining the polarisation change caused by the device under test (DUT). A transmitting antenna (TX) produces a plane wave of known polarisation. A calibrated polarisation rotator is adjusted until the fixed receiving antenna (RX) is recording the maximum field amplitude. The setting of the known device corresponds to the polarisation changed caused by the DUT.

thermore, it requires neither the rotating receiver, nor a known device for comparison.

The ability to completely determine the polarisation change of the VADR device is not compromised either. The key properties of any retarder are its retardance Γ, equal to the difference in phase between the TE and TM transmission coefficients, and its TE and TM insertion losses, given directly from the magnitudes of the TE and TM transmission coefficients, respectively. With this information provided by the measurements, the behaviour of the VADR for any given incident wave may be calculated using either the Jones calculus or the Stokes parameters (see section A.2).

This experimental design calls for the test setup to be characterised, and then for the VADR to be measured at all separations for the TE and then the TM polarisation as shown in Fig. 5.4. The existing W-band (70 – 110 GHz) measurement equipment will be described in detail in due course, but for now it is necessary to mention that it was equipped with pyramidal horn antennae, which emitted vertically-oriented, linearly polarised radiation, indicated by the arrow marked **E** in Fig. 5.4. In the first set of measurements, the VADR grooves are mounted horizontally. Since the grating vector **k** is defined perpendicular to the grooves, the first set of measurements determines the TM response. The complex transmission coefficient, in the form of S-parameters, is recorded for all the positional settings 0-3000μm separation, across the W-band frequencies 70 - 110GHz, including the design frequency of 100GHz. In the second set, the VADR device is rotated by 90° so that the TE is measured.

It should be noted that no polarisation change is possible when the VADR device is illuminated by a purely TE or TM wave, because both must be present for birefringence to occur. This is important because it means that the two horn antenna need only be identically oriented and do not need to be adjusted during the experiment. When used in an application with a transmitting pyramidal horn antenna, the VADR device would simply be oriented at 45° so that it would appear to be illuminated by both TE and

TM and thus able to produce birefringence.

As an additional check on the measurements, two plain silicon plates have been measured. The two plates would form an etalon (Fabry-Pérot cavity) and this should be evident from Fabry-Pérot oscillations in the measurements. However, plain silicon is not birefringent so it should be possible to measure the etalon, rotate each plain silicon plate by 90° and obtain the same results upon re-measurement. Also, if the peak transmission magnitude is less than 0 dB then it will indicate that there is loss in the silicon plates. Some loss is expected because the silicon has low resistivity.

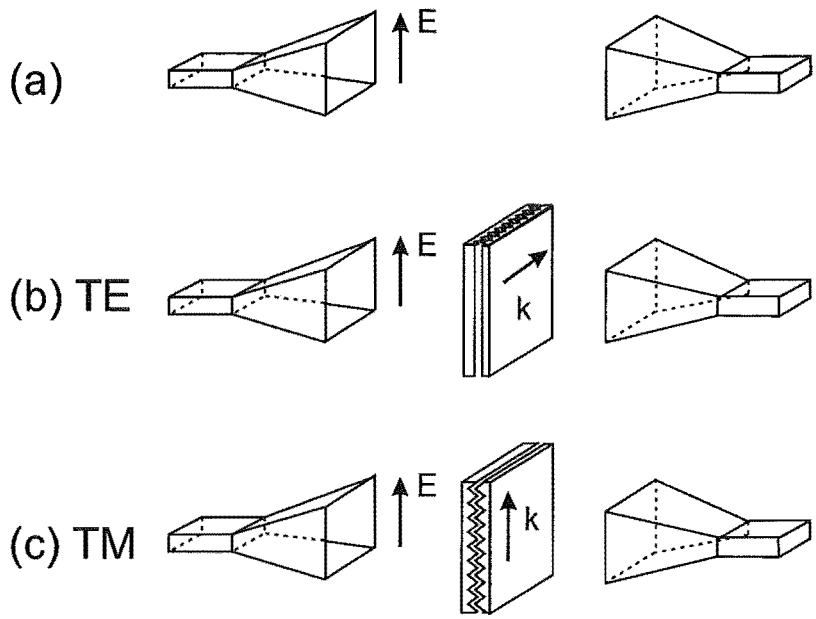


Figure 5.4 Proposed experimental setup, showing the two horn antennae, and the VADR device mounted between them. The transmitted electric field **E** is oriented vertically. The complex S-parameters are measured for: (a) the measurement setup without the VADR; (b) the TE polarisation; (c) the TM polarisation.

5.2.3.1 S parameters

The transmission and reflection coefficients of linear microwave devices are often presented in terms of the complex scattering-parameters (S-parameters), which give both magnitude and phase information [159]. It is the phase information that is of primary interest in the VADR measurements, because this quantifies the VADR device's key property, the retardance Γ. However, the magnitude information is important as well, because it provides a figure for the insertion loss and an indication of the error introduced into the polarisation tilt.

The S-parameters for a two port network are defined as

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \tag{5.1}$$

where S_{11} is the input reflection coefficient, S_{21} forward transmission coefficient, S_{12} reverse transmission coefficient, and S_{22} output reflection coefficient. Each parameter has a magnitude expressed in decibels (dB) or as a ratio, and a phase in degrees or radians. The VADR devices are the same in both forward and reverse, so it is only S_{11} and S_{21}, or S_{22} and S_{12} that need to be considered. The phase difference and the insertion loss are determined from S_{21}, and S_{11} gives the fraction of the loss due to reflection. The S-parameters do not contain any information about the polarisation state, and this must be recorded separately (usually by hand).

5.3 EQUIPMENT DESCRIPTION

This section describes the equipment used to measure the VADR device's performance. The measurements were conducted in the Ultrafast laboratory at the University of Glasgow, using the existing W-band free-space measurement setup. Due to the strict requirements for a linearly polarised beam, the horn antenna were theoretically analysed to determine whether additional polarisation control was required. Special fixtures were required for mounting the VADR plates, and controlling their separation, and these are also described.

5.3.1 W-band Free-Space Measurement Equipment

The measurement system comprised a vector network analyser (VNA), a synthesised frequency source, two transmission/reflection (Tx/Rx) measurement modules and their associated test platform, and two pyramidal horns. A block diagram showing how these items are connected up for free-space measurement is shown in Fig. 5.5, where the device under test (DUT) is not specified. The VNA controls the excitation frequency (via the frequency synthesiser), and measures the signals received by the antennae. The Tx/Rx modules up-convert the synthesised sweeper's signal into the W-band for transmission, and down-convert the received signals for transmission to the VNA via coaxial cable.

The Wiltron 360B VNA, shown in the photograph of Fig. 5.6, can measure all four S-parameters of a two port network (S_{11}, S_{21}, S_{12}, S_{22}). Measurements are displayed on the integral screen, and may be transmitted over a General Purpose Interface Bus (GPIB) to an accompanying Intel x86 architecture PC, running Microsoft Windows and Cascade Microtech's program Win-Cal version 2.2 [160]. Win-Cal allows the measured S-parameters to be displayed on screen in either polar or rectangular co-ordinates, and they may also be saved to disk for further analysis.

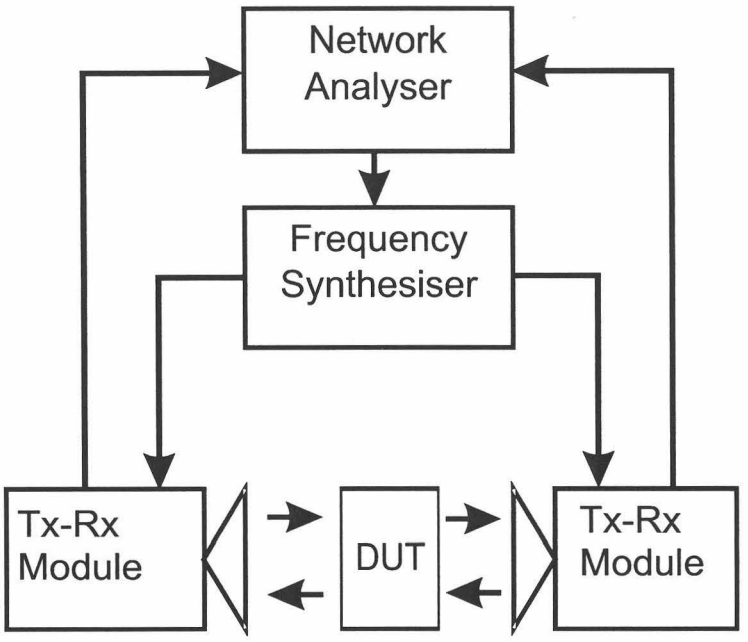


Figure 5.5 Block diagram of the W-band equipment interconnections in free-space measurement configuration

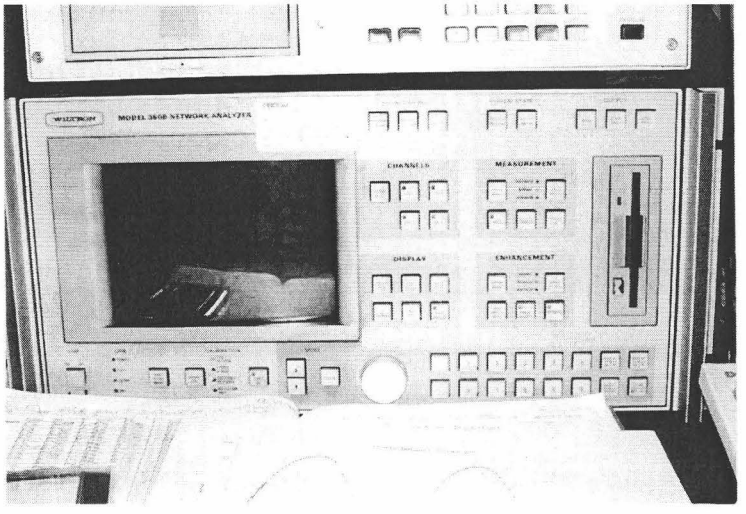


Figure 5.6 Photograph of the front panel of the Wiltron 360B vector network analyser.

The Wiltron 68187B synthesised sweep generator provides a continuous wave source at -15dBm – +12dBm over the frequency range 10MHz – 60GHz. For the VADR experiments the source was operated at a power setting of -15dBm, as this was the standard laboratory procedure. In order to produce the W-band frequencies, 67 – 110 GHz, two Wiltron SM4873 transmission-reflection modules are used to up-convert the signal provided by the frequency synthesiser. These transmission-reflection modules also contain a dual directional coupler so that signals can be both transmitted and received, as the name suggests, avoiding the need to manually reverse the test set to measure all four S-parameters. The modules also down-convert the signals, for transmission back to the VNA. The VNA is programmed to display the actual test frequency frequency on screen, so the process is transparent to the user.

5.3.2 Pyramidal Horn Antenna

The TX/RX modules were equipped with a pair of waveguide-fed WR-10 pyramidal horn antennae for free-space measurements. The horns are constructed from high grade bronze, and powder coated in black on the outside. The horns are shown in place, in Fig. 5.7.

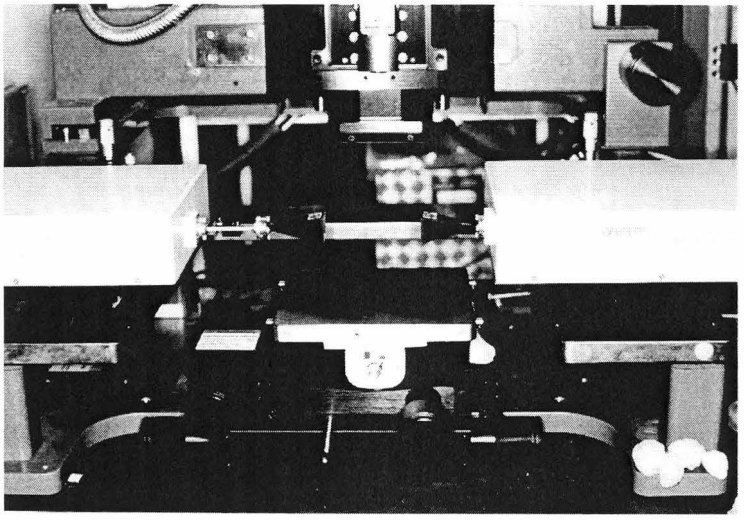


Figure 5.7 Photograph of the pyramidal horns mounted on the transmission-reflection units.

An important requirement of the experimental design is that the transmitting and receiving antennae are strictly linearly polarised. This is critical, because otherwise, it would be impossible to independently measure the TE and TM transmission coefficients. Fortunately, the present experiment is completely safe in this regard since it is has long been known that a pyramidal horn antenna, excited by the dominant mode of a rectangular feed waveguide, emits a "strictly linear polarization" [161].

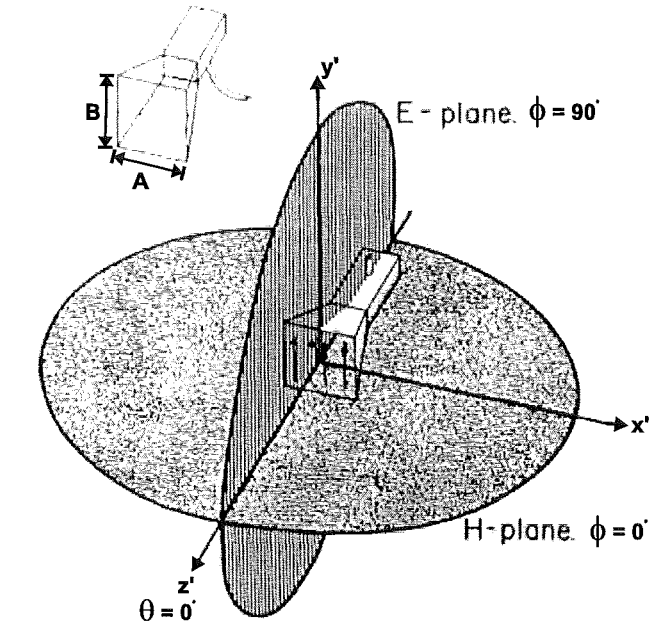
An experiment using the same horns and feed waveguide was conducted to confirm that the horn antennae were indeed linearly polarised. In order to allow the horn antennae to rotate, a Gunn oscillator source (94 GHz) was used along with a spectrum analyser having an external mixer. The noise floor of this set-up was -32.5 dB below the maximum output of the Gunn oscillator. The restricted dynamic range is a consequence of the high conversion loss (30 dB) in the external mixer. As a result, the best self-extinction ratio that could be measured was 32.5 dB (i.e. limited by the noise floor), which indicates that less than 0.1% of the total power is present in the undesired (cross-polarised) electric field component. This is acceptable for the present experiment.

The field pattern of the horn antenna was calculated theoretically, and the results are plotted in Fig.5.8. The orientation of the E- and H-plane field patterns of a horn antenna are defined in Fig. 5.8(a). The field patterns themselves depend on the horn dimensions, and may be calculated analytically according to the method in Balanis' text [162]. It is important to know the width of the beam that is encompassed by the VADR device, so that the it can be gauged whether there might be field leakage around the edges of the device. The beam patterns calculated for the present horn antennae are plotted in Fig. 5.8(b).

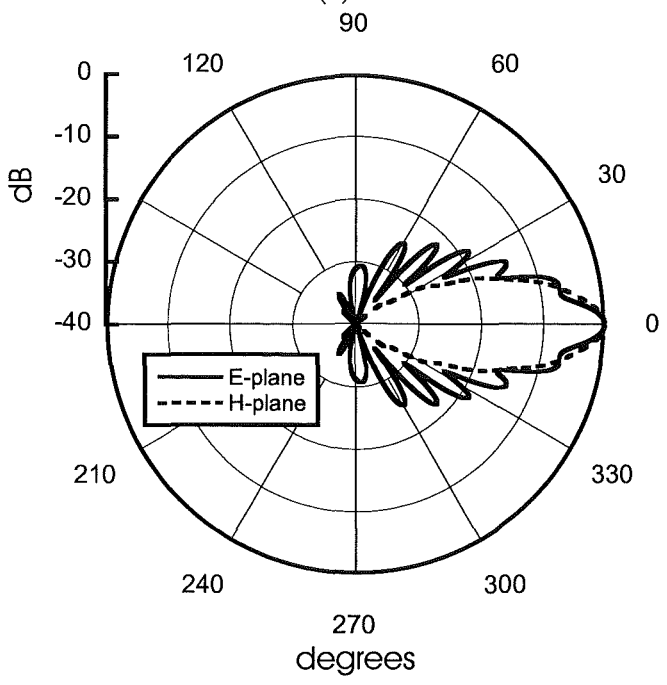
Given the intended 75 mm separation between the VADR horns, the exposed 20 mm by 20 mm area of the VADR device is predicted to encompass the beam out to -11 dB in the E-plane and -13 dB in the H-plane. The difference in values arises because the beam is slightly wider in the E-plane. This is acceptable for the present experiment, although ideally, the VADR would have encompassed the beam out to - 30dB (0.1% power) in both E- and H-planes. However this would have required a device that was in excess of 200 mm across - this is impractically large. Alternatively, the beam could be made narrower by using a focusing lens. Unfortunately, suitable lenses were not available commercially for the existing horns. Custom construction of suitable lenses was outside the scope of the present project. Therefore, it was decided that apertures constructed from absorbing foam would be placed on either side of the VADR device to reduce field leakage.

5.3.3 VADR-Specific Fixtures

Special fixtures were required to mount the VADR plates and control their separation. While only one axis of control (the plate separation) was required during the experiment it was also important that the plates could be aligned to each other beforehand. The fixtures were designed before it was possible to inspect the measurement equipment directly, so the dimensions were estimated from photographs [163]. The test platform between the two antenna was estimated to be 150mm by 150mm laterally, and the space between the two horn antenna appeared to be approximately 60 mm although it was later found this distance could be increased further to 75 mm.



(a)



(b)

Figure 5.8 The E- and H-plane field patterns calculated for the WR-10 horns: (a) definition of E- and H-planes; (b) calculated field patterns plotted against angle and log magnitude. The horn's open mouth is directed towards 0°.

The two VADR plates were mounted on separate fibreglass apertures, one of which was actuated by a translation stage, as shown in the plan view of the equipment layout in Fig. 5.9.

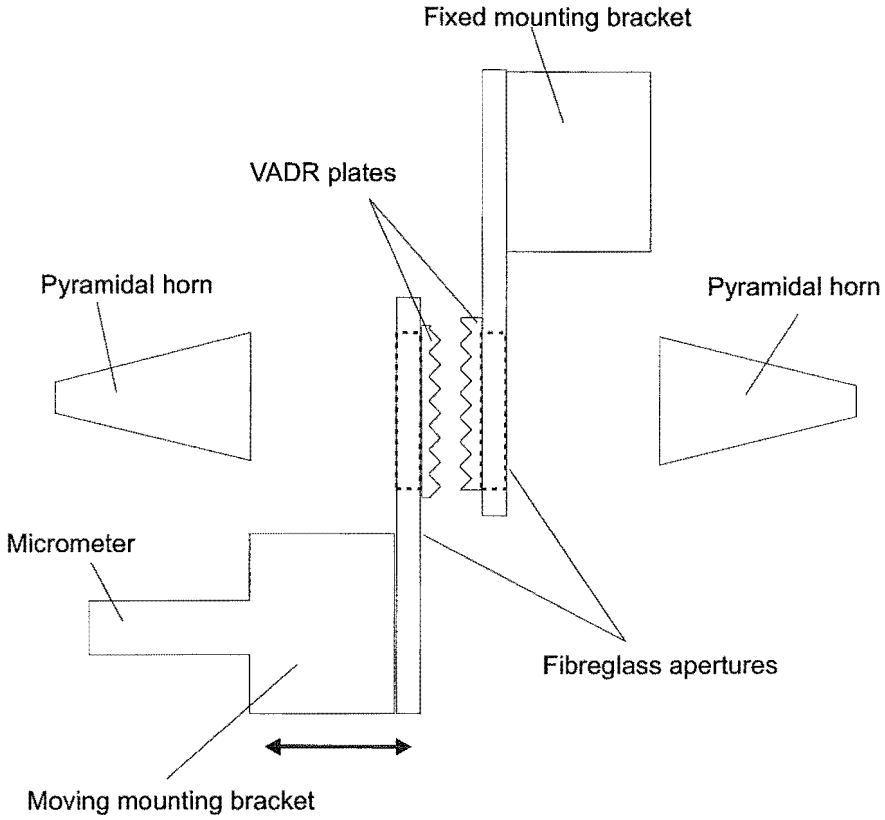


Figure 5.9 Schematic of the layout of the VADR test fixtures

The VADR plates were mounted over fibreglass apertures, drawn schematically in Fig. 5.10. The dimensions of the aperture were 20mm by 20mm, so that the larger (22 mm) side of the VADR plates overlapped enough to affix them to the apertures with double-sided adhesive foam while, the four screw holes in the plate correspond to the hole grid-pattern on the face of the two 90° angle mounts (New Focus 9039), one for each plate. Fibreglass was chosen because it was less reflective at these frequencies than metal, yet was still stiff enough for accurate dimensional control. This was important because the plate separation was to be controlled by positioning the apertures.

The angle bracket associated with the moving plate was mounted on a New Focus pint-sized 9065-X translation stage, which was actuated by a micrometer of resolution 10μm and travel 13mm (New Focus 9352). The angle bracket associated with the fixed plate, was mounted on a stack consisting of a baseplate (New Focus 9031) and a 0.5" modular riser (New Focus 9036) so that it was mounted at the same height

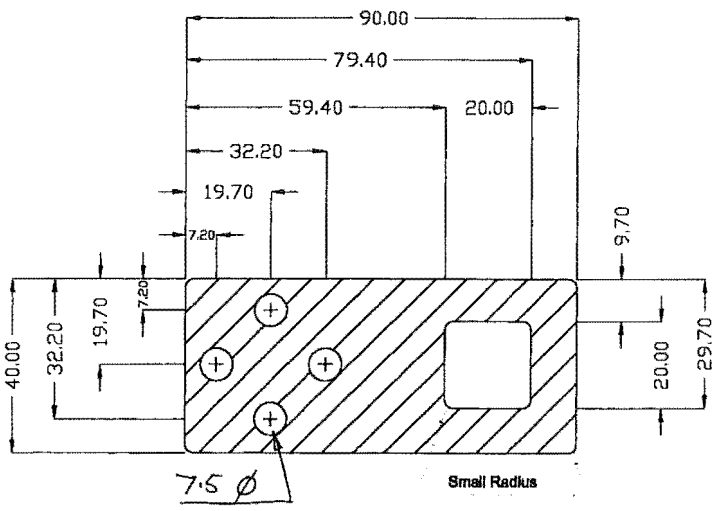


Figure 5.10 Engineering drawing for the fibreglass apertures used to mount the VADR plates. The VADR plate sits across the square hole, while the four round holes are for mounting the aperture to the rest of the fixtures.

as the moving angle bracket. A custom baseplate was constructed from 10mm thick aluminium, to match the platform of the test equipment. Such a thickness was not required to handle any large mechanical forces, but rather to assist with temperature stability of the system. A true section of aluminium was selected, rather than use milling to re-establish planarity. Connections between the New Focus components, and to the baseplate, was by means of ez-Trac (New Focus 9032). All screws and screw holes used an M6 thread, and the supplied screws (New Focus 9024) were complemented by some longer M6 screws for use in mounting the fibreglass aperture. Small rubber O-rings were used like washers, to provide better control over the amount of grip on the fibreglass apertures, without over-stressing the screw threads in the angle bracket. An additional bar of aluminium was used to true the two angle brackets, and the line of motion of the micrometer stage. The bar was affixed to the face of one bracket, and the other bracket was rotated until it aligned with the bar. The face of the angle bracket on the translation stage was carefully adjusted until it was perpendicular to the line of motion. A photograph of the VADR mounting fixtures may be seen in the next section.

5.4 EXPERIMENTAL PROCEDURE

This section details aspects of the experimental procedure that fall outside the scope of the experimental design as described in section 5.2. The first part of the section describes how the VADR plates were mounted on the apertures for the experiment. The second part describes the procedure for taking measurements and moving the

translation stage.

5.4.1 Handling precautions

Several precautions were observed to minimise disturbances to the measurement equipment, in order to ensure repeatability of the measurements. Firstly, the equipment was never switched off. Heating effects in the microwave components cause the couplers and waveguides to expand, so the temperature must be allowed to stabilise before measurements can be taken, so that results are repeatable in the future. Should the processor in the vector network analyser hang, it was hot-reset rather than having the power supply cycled. For the same reason, when wafer probe station was converted for free space measurement by replacing the wafer probes with pyramidal horns, the system was left to thermally stabilise overnight where possible.

Further precautions were observed with respect to the handling of the waveguide test standards and the pyramidal horns. In no circumstances were the items to be treated roughly, and particular care was taken to protect the mating faces from scratches. The WR-10 waveguide components are aligned with the aid of four pins, and held in place by four screws. A pre-set torque wrench, supplied by the manufacturer, was used to evenly tension the screws to 6 cNm. Fingerprints on any mating surfaces were cleaned with acetone.

5.4.2 Calibration Procedure

Good calibration is an essential ingredient in obtaining useful results. Unfortunately, it was not possible to include the horns in the calibration because without lenses they cannot be focused onto test standards. Therefore, a waveguide calibration was conducted that extended to the flange connecting to the horn. This was not inconvenient because it was still necessary to de-embed the test fixtures from the measurements in any case, and the behaviour of the horns was simply another aspect of the test fixtures to be de-embedded.

There are several standard calibration procedures for vector network analysers, and they are usually referred to by the three or four letter acronym, representing the type of calibration standards employed such as through-lines (L), short circuits (R for reflection) and loads (M for match). For two port networks, such as the free space measurement system, the preferred method is the more accurate Line Reflect Line (LRL). The Wiltron 360B also supports the easier-to-use Line Reflect Match (LRM) and Through Reflect Match (TRM) methods, however the LRL calibration method was used exclusively for these experiments since that was the procedure already in place.

An exploded-view schematic of the calibration standards and the calibration procedure is shown in Fig. 5.11. A metrology grade bronze waveguide calibration kit,

supplied by Flann [164], provided the quarter-wave through-line and short-circuit. The calibration procedure was conducted as follows. First, the two measurement modules were connected together with a 35mm length of WR-10 waveguide. This length of waveguide remains attached to module one throughout the measurements. The measurement at module one is referenced to the far end of this waveguide. Second, a 1.0860mm thick through-line standard (nominally $\frac{\lambda}{4}$ at 69GHz) was inserted between module two and the waveguide. This was the second through-line measurement, and it is performed out of order with respect to the acronym describing this method (LRL). Thirdly, the through-line was replaced by a short-circuit (reflective) standard, and measured first on port one. The short-circuit was then transferred to port two, where it was remeasured.

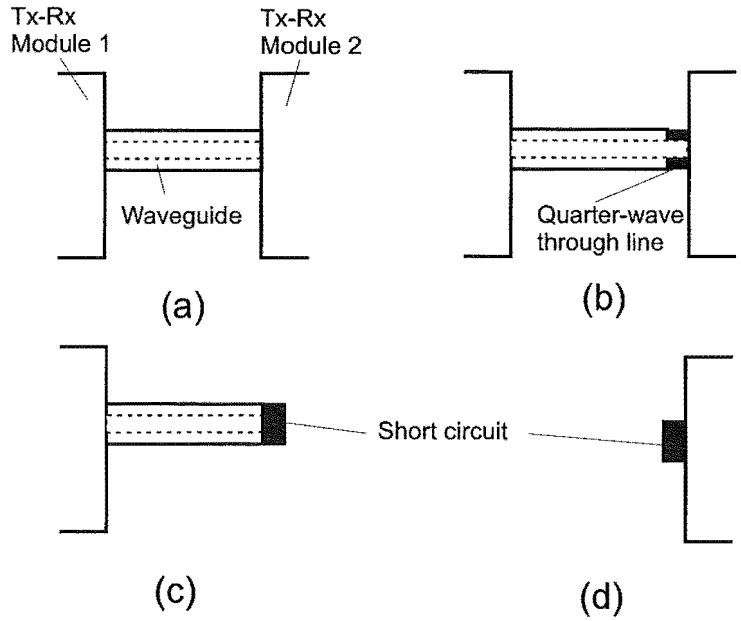


Figure 5.11 Schematic of the calibration standards and the calibration procedure: (a) Through-line measurement, (b) second through line measurement - with extra quarter-wavelength waveguide standard, (c) short-circuit port one, (d) short-circuit port two.

Once the calibration procedure was complete, the calibration data could be saved to PC-format 3.5-inch floppy disk, for later retrieval e.g. if the machine needed to be reset. Thereafter, the calibration was automatically applied to all measurements, including those presented on the VNA screen and those sent over the GPIB bus to be saved for later analysis. This feature could be turned off if desired, and raw data collected, however this was not desired for the present experiment. It was carefully checked throughout that the calibration remained automatically applied to the measurements, by way of the "calibration setting" indicator light on the front panel. It should also be noted that throughout the calibration and the measurements, the system was set to

measure each point in the frequency sweep 256 times, and record the average value.

5.4.3 Mounting the VADR plates

It is important to have correctly aligned VADR plates, as illustrated in Fig. 5.12(a). The grooves must be centred with respect to each other so that they can interlock freely and easily. If the plates are offset laterally as in Fig. 5.12(b), then the forces that the plates exert on one another can disturb the accuracy of the plate positioning. If the plates are not parallel to each other as in Fig. 5.12(c), then the two plates cannot fully interlock.

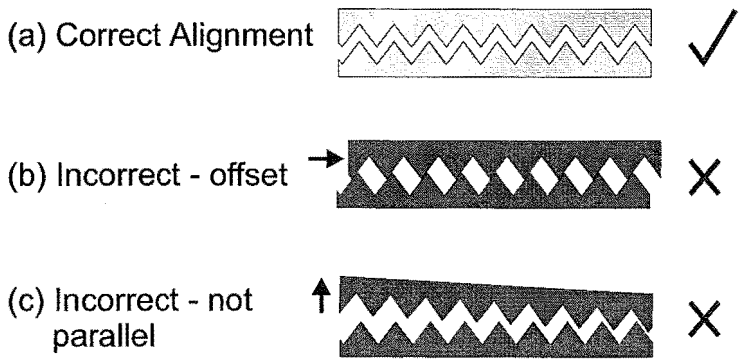


Figure 5.12 The VADR grooves must be exactly aligned when mounted for testing (a). Two examples of incorrect alignment are provided for comparison, a lateral offset (b), or the plate faces not being parallel to each other (c).

The plates were affixed to the fibreglass apertures with double-sided adhesive foam, 1.6mm thick. The foam was purchased from a stationary store, and was of generic composition. The double sided foam provided mechanical cushioning at full interlock, and allowed the plate alignment to be finely adjusted at mounting time without risking plate breakage. It is worth noting that the glue was quite strong and formed a good bond instantly, when gentle pressure was applied. It was necessary to use acetone to dissolve the glue bond before the plates could be removed after the experiment.

The following steps were used to mount the VADR plates.

Step 1

Make sure that the apertures are aligned and parallel, using the aluminium bar to true up the fixed and moving angle brackets. Clamp the mounting stage to a vertical bench support (this is what was handy at the time).

Step 2

Affix the first plate to the lower aperture, using double sided foam, with the grooves facing up.

Step 3

Affix the double sided foam to the upper aperture. Place the second plate on top of the first plate, so that the grooves interlock, as shown in the photograph of Fig. 5.13. The plates are largely self-aligning. Slowly lower the second aperture using the micrometer, until the foam pads adhere to the second plate.

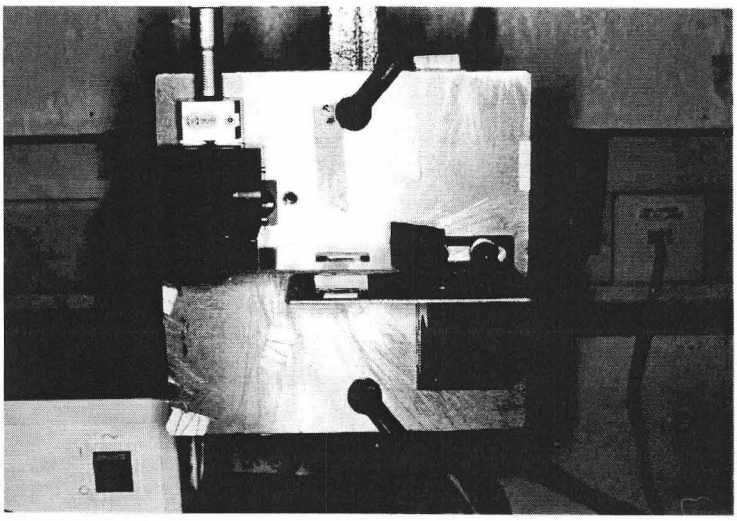


Figure 5.13 The VADR plates are aligned before the second plate is affixed to its aperture.

Step 4

Inspect the plate alignment using a microscope. Manipulate the plates until a good alignment is achieved. Absorbing foam apertures are affixed with double-sided tape to the outer sides of the fibreglass apertures, to reduce field leakage. The fixture is let to rest, in case the foam pads creep, and then the alignment is rechecked. When the alignment is accurate and stable, the stage is unclamped from the vertical bench support and transferred to the W-band test station platform.

The position of the test fixture is located with masking tape and pencil marks, and fixed with two clamps. Further absorbing foam is placed over all metal surfaces in the test fixture, particularly immediately beneath the device, as this halfway point between the two horns is a likely source of error from stray reflections if left uncovered. The entire test fixture is covered over with absorbing foam as an extra precaution against stray reflections.

The transmit and receive horns are widely separated, by approximately 25λ, to ensure a plane wavefront at the device. The electric field emitted from the horns is linearly polarised. The experimental design called for independent measurement of transmission coefficients for TE and TM electric fields. The VADR was first mounted with grating wavevector $\mathbf{k}$ parallel to the electric field, for the TM measurement, where

k is defined in Fig. 5.4. The plate separation distance was set at one free space wavelength (3mm) and reduced in 60μm steps until the grooves just reached the interlock region (<350μm). Positional errors from backlash in the micrometer were eliminated by always moving the micrometer in one direction only. Since the sensitivity of the device was greater in the interlock region, the separation distance was reduced with a step size of 20μm until full interlock was reached (0mm). The plates were re-mounted with **k** perpendicular to the electric field for the TE measurement. The measurement steps were repeated. At each separation, the S-parameters were recorded for 256 frequencies between 70 and 110GHz.

5.5 RESULTS AT 100GHZ

This section presents the measurement results of the VADR device in the W-band, but particularly at the design frequency of 100 GHz. First, however, some results pertaining to validation of the measurement setup are presented.

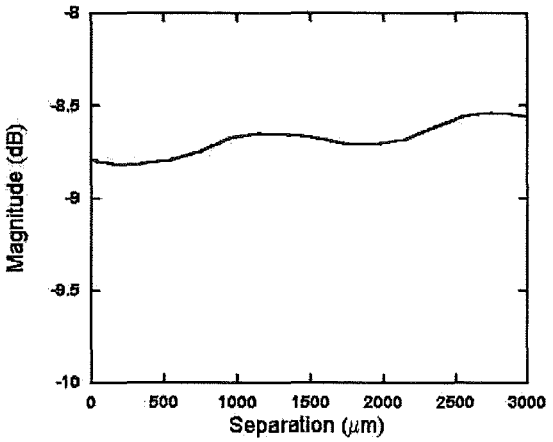
5.5.1 Measurement precision

The measurement precision was estimated in the following way. After calibration, the horns were spaced widely apart (75 mm)and the region in between was surrounded on all sides by absorbing foam to create a crude pseudo-anechoic chamber. The phase and magnitude of S_{21} was monitored at 100 GHz over several minutes, while sweeping over the whole W-band range. This gave more than ten separate readings. Even at low levels of averaging (four points per measurement), the variations in the readings were better than $\pm$0.01 dB and $\pm$0.1°.

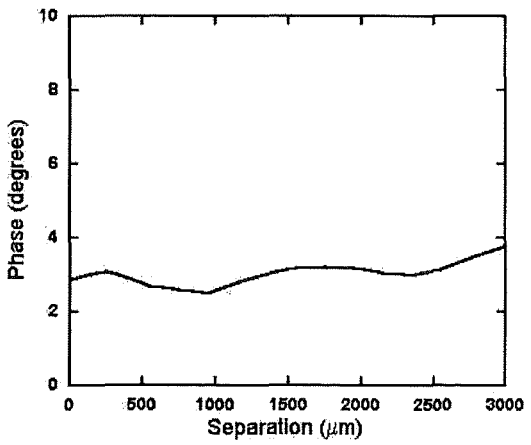
5.5.2 Apertures only

In order to quantify the insertion loss due to the mounting stage and apertures, these were characterised over the complete range of holder separations that would be employed when the VADR device was measured. However, for this experiment, the VADR plates themselves were deliberately omitted, so that the insertion loss due to the test fixtures could be separately quantified. The S_{21} responses plotted in Fig. 5.14 show little change in the loss or phase shift as the apertures are separated. The insertion loss was -8.7$\pm$0.1 dB across all separations. In order to accurately quantify the insertion loss of the VADR device, the exact value of the test fixture insertion loss may be subtracted from the combined insertion loss of the VADR device and the test fixture. The phase was 3.2$\pm$0.2° across all separations. The absolute phase shift caused by the mounting plates is unimportant, however the variation in phase shift indicates that the phase measurements would be no more accurate than $\pm$0.2°. These variations

± 0.1 dB and $\pm 0.2°$ are taken as the experimental uncertainties, since it is a more conservative approach than taking the smaller errors obtained from a static setup as in section 5.5.1. Even for these uncertainties, error bars would be too small to see, so they are not plotted in any of the following graphs of measured results.



(a)



(b)

Figure 5.14 Measured complex transmission coefficient (S_21) for the test fixtures alone (no VADR) at 100 GHz: (a) magnitude (a), (b) phase shift.

5.5.3 Plain silicon plates

A control experiment was conducted with two silicon plates, of the same dimensions as the VADR plate, and cut from the same batch of silicon wafers. The measured complex transmission coefficient is plotted in Fig. 5.15. The phase is raw data, but the magnitude has the test fixture's loss removed. The expected etalon effect is clearly evident. The loss varies from -8 – -15 dB, and the phase varies over a range of 18° (the absolute figures are unimportant, but in this case the phase varies from 49 – 67°). There is no change in the response when the samples are rotated - which is as you would expect because plain silicon is not birefringent. While this seems like an obvious experiment, it is useful to show that plain silicon is not birefringent.

Also, this experiment indicates that the silicon substrate is appreciably lossy, since the magnitude peaks well below 0 dB. In the following chapter, the loss will be estimated from this measurement through comparison with simulations of the etalon.

5.5.4 VADR device

The VADR plates were then inserted and measured. The complex transmission coefficients for the design frequency of 100GHz are plotted in Fig. 5.16, with TE magnitude in Fig. 5.16(a), TM magnitude in Fig. 5.16(b), TE phase in Fig. 5.16(c) and TM phase in Fig. 5.16(d). Note that the measured phase is raw data, but the measured magnitude data has the test fixture's insertion loss (8.7 ±0.1 dB) removed. The measured curves show two distinct regions of operation, one corresponding to interlocked plates (0 – 350μm), the other separated plates (>350μm). In the separation region, Fabry-Perot cavity oscillations are evident, with a period of half the free-space wavelength. The magnitude responses, Figs. 5.16(a,b), show a distinct asymmetry due to the birefringent grooves. Over the entire operating range, the insertion loss oscillates from 4 – 7dB for TE and 3 – 11dB for TM radiation. The phase response also exhibits asymmetry, oscillating from $-20° - -30°$ for TE, and 0 – 50° for TM radiation.

5.5.5 Birefringence

The difference in the phase response between TE and TM polarisations is the birefringence property that is desired, and this is plotted in Fig. 5.17. This gives a measure of the device's retardance as a function of the plate separation, $\Gamma(s)$. The maximum retardance of 74° was measured at a separation of 830μm, which is approaching a quarter wave. There are two distinct areas of quasi-linear variation in $\Gamma(s)$. The first is in the interlock region, where separation distance is less than the groove depth (0 – 350μm). The retardance changes from 22° to 34°, for separation distances 0 – 100μm, yielding a sensitivity of 120°/mm. A 0° retardance was not observed due to mechanical imperfections inhibiting full interlock (scanning electron microscope images in Fig. 4.17

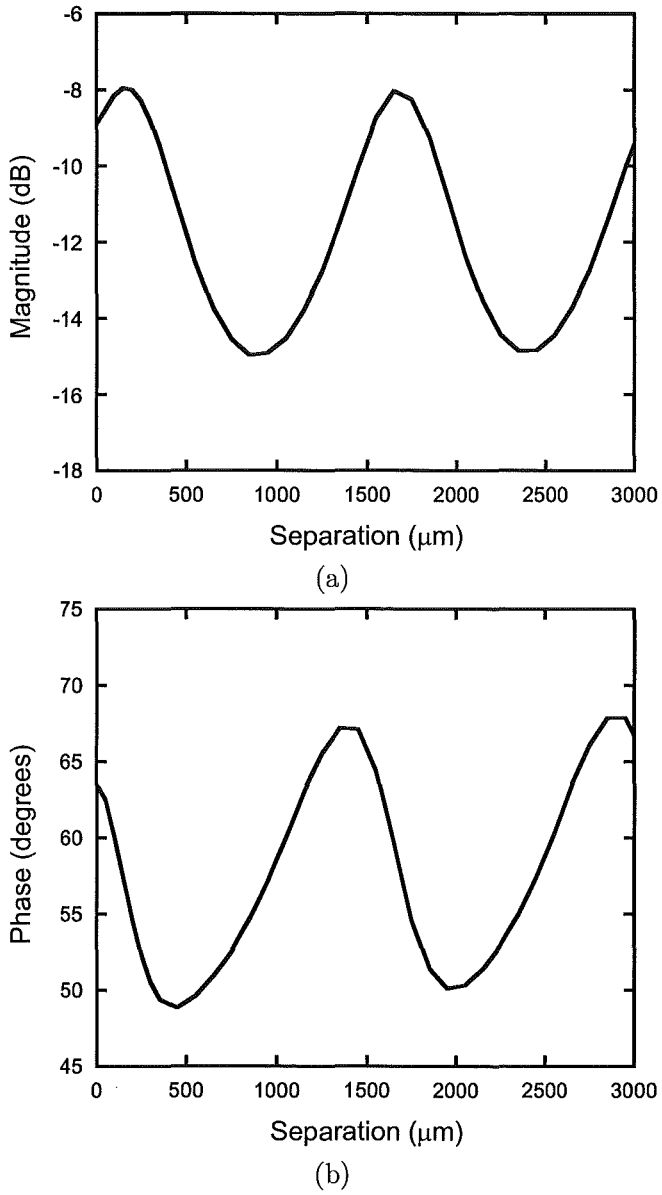


Figure 5.15 Measured complex transmission coefficient (S_{21}) for two plain silicon plates at 100 GHz: (a) magnitude (a), (b) phase shift. Note that the phase shift is raw data but the magnitude has the test fixture's insertion loss removed.

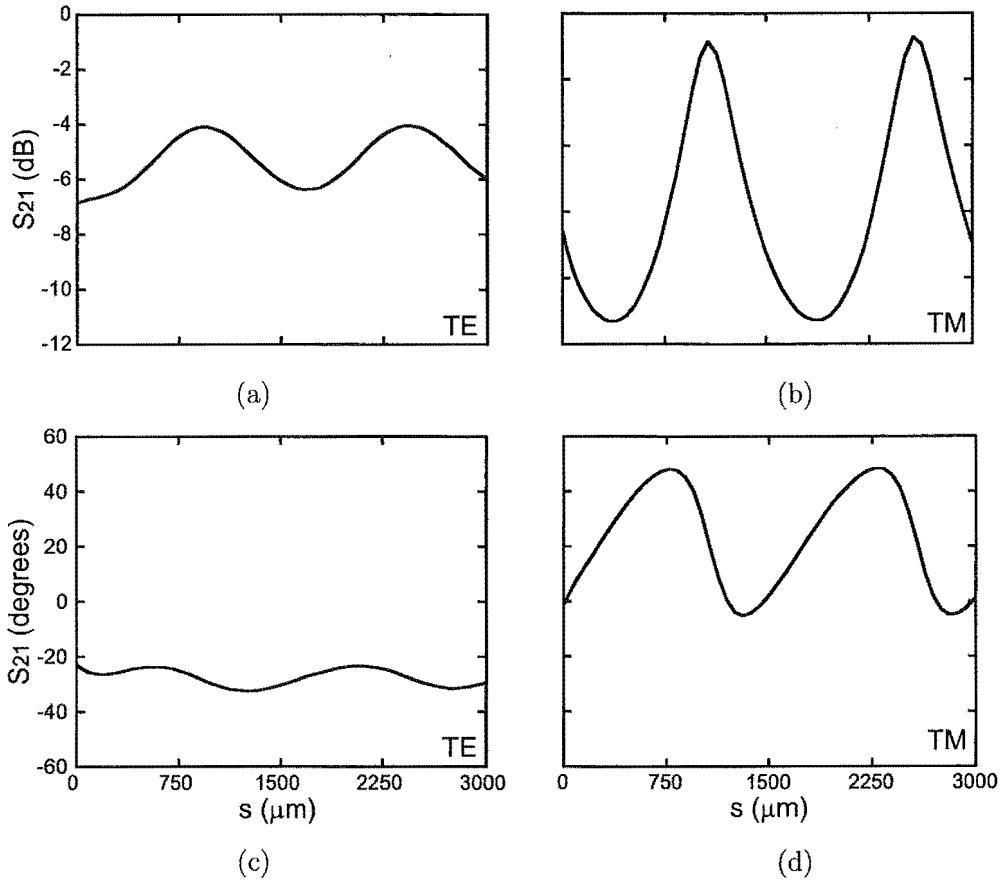


Figure 5.16 The measured, complex transmission coefficients of VADR variable polarisation compensator at 100 GHz for incident TE and TM linearly polarised radiation. (a) TE magnitude, (b) TM magnitude, (c) TE phase, and (d) TM phase. Note that the phase shift is raw data but the magnitude has the test fixture's insertion loss removed.

of a device in full interlock show a plate separation of 20μm, resulting from surface irregularities). This could be compensated for by adding an opposing, fixed, birefringent plate to the outer surface of the device; for example a grating could be etched into the outside of one of the plates, perpendicular to the V-grooves.

A second region of quasi-linear variation occurs for separations of $1000 - 1300\mu$m, giving a relative change in retardance of 36°, and sensitivity of $-120°$/mm. Note that this is in the separation region, where the birefringent-mirror Fabry-Perot behaviour dominates.

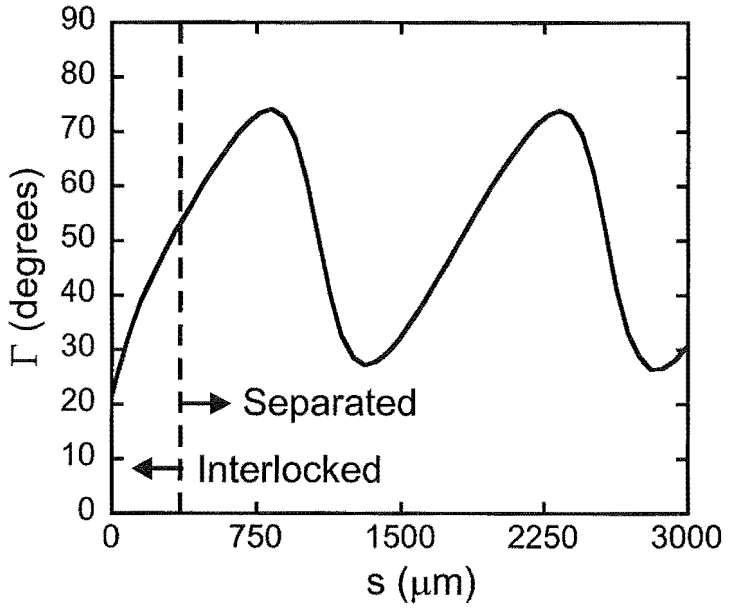


Figure 5.17 VADR retardance as a function of plate separation at 100 GHz, $\Gamma(s)$, from measured data.

In the rigorous numerical results obtained by Cumming and Blaikie [65] in their proposal, for a 1 THz device, the peak retardance of 60° was obtained at 75μm - the equivalent to 750μm here. Thus, the experimental device has demonstrated a greater-than-predicted retardance. However, the insertion loss is also greater than predicted, at 3.0 dB for TE and 3.2 dB for TM. Additionally, the qualitative shape of the phase difference between TE and TM (the retardance Γ) is different. In the simulations, the phase difference has approximately the same slope either side of the peaks, whereas for the measured results, the slope is noticeably steeper on the side of larger separations. These observations could be attributable to the lack of anti-reflection coatings and further investigation is warranted. This is pursued in the following chapter.

5.5.6 Frequency dependence

The frequency-dependence of the device has also been characterized by measurement. The retardances at the upper (110 GHz) and lower (70 GHz) ends of the measurement range are shown in Fig. 5.18. For the 110 GHz trace, there is shown a full period of the separation-region oscillation. The period of the oscillation corresponds to half the free space wavelength, $\lambda_0/2 = 1.35$ mm, as would be expected from a Fabry-Pérot cavity. The effective groove depth (the groove depth divided by the free space wavelength) decreases at lower frequencies, lowering the effective path difference between the TE and TM polarizations. Consequently, there is a reduction in the peak retardance and sensitivity at 70 GHz, and the position of peak retardance moves to greater separation distances. It is seen from these measurements that the device operation is dispersive, but it operates as a variable polarization compensator over a wide bandwidth, and the sensitivity increases with frequency.

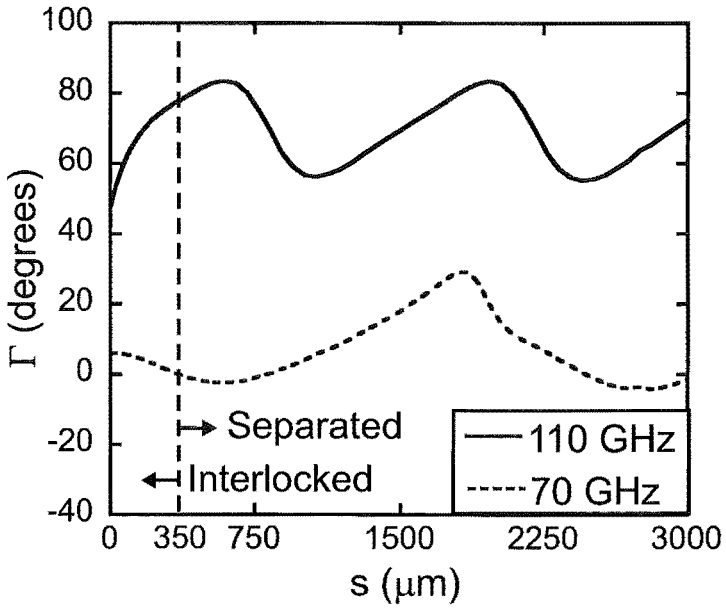


Figure 5.18 VADR retardance $\Gamma(s)$ at the upper and lower ends of the measurement range, 70 and 110 GHz.

5.6 SUMMARY

The VADR device was tested at W-band (67-110GHz) frequencies using a vector network analyser in free-space configuration. Pyramidal horns were used to launch and detect the radiation. The author designed a mounting stage to hold and actuate the VADR plates, using a combination of the off-the-shelf optical components, custom fibreglass apertures and a custom aluminium baseplate. The experimental design specified

that the device's complex transmission coefficients for TE and TM radiation be separately determined, and the difference in phase shift calculated, to give the retardance (polarisation shift) of the device. The maximum retardance of the device was measured to be 74°, with a maximum sensitivity of 120°/mm in the first 100μm of the interlock region. The shape of the response differs from the behaviour predicted in the device proposal. The analysis in the original proposal does not take reflections from the back walls of the plates into consideration, and this may be cause of the discrepancy. Further investigation of the modelling of this device is warranted.

Chapter 6

VADR SIMULATION

6.1 INTRODUCTION

Now that the complex transmission coefficients of the VADR device have been experimentally measured, and shown to agree only approximately with the existing simulations, further work on the simulations is warranted. Good agreement between simulations and measurements is desirable because it indicates that the operation of the VADR device is well understood. It is expected that the present discrepancies arise from the omission of anti-reflection (AR) surfaces from the rear surfaces of the measured VADR. The AR coatings were assumed to be present in the original simulations, but were omitted from the measured devices because a suitable process for creating them had not been developed nor their performance characterised. Thus, it is desired in this chapter to simulate a device that has no AR coatings. Since the omission of the AR coatings gives rise to reflections at the rear surface of the device, the effect of these now needs to be included in the simulation model. This complicates the modelling task compared to the original [65], which did not model the AR coatings directly, but achieved the same effect by omitting the air-silicon interface at the outside surfaces of the device. Hence, more boundary conditions are required to be added; in this case, the multiple-multi-pole (MMP) numerical tool used by Cumming and Blaikie may not provide the most straightforward modelling approach.

In this chapter, several simulation techniques are reviewed, including a detailed comparison with calculated results and effective medium theory results for subwavelength 1-D grating. The VADR device is then analysed with the selected rigorous coupled-wave analysis (RCWA) technique.

6.2 REVIEW OF ANALYSIS AND SIMULATION TECHNIQUES

Despite the elegance of Maxwell's Equations, often a significant effort is required to solve them for the electromagnetic behaviour of a structure. Unfortunately, even with the best choices of approximations, they are difficult or impossible to solve analytically for any but the simplest structures. Fortunately, with the wide availability of powerful

computers, these restrictions need no longer apply, and the behaviour of structures with complicated geometries may be solved for using numerical techniques.

In general, the computer-based numerical tools (codes) are developed with the minimum of assumptions and approximations, particularly with regard to actual problem geometries, so that they may be applicable to a broad range of problems. Since electromagnetic codes have been under development for around thirty years, it is now often found that there already exists one or more that appear suitable for a given problem. Then it becomes necessary to weigh the suitability of the various numerical algorithms to your problem, but also to choose whether to use an existing implementation or write your own.

Where software development is not the main theme of a study, and if a suitable implementation is available, then it is preferable to choose existing code over writing your own. This is because creating numerical solvers is a non-trivial exercise, and many aspects of making suitable electromagnetics code are far removed from the field of electromagnetics and are fields of study in their own right. Thus, in the present case, the choice of numerical technique depends on the capabilities of available software packages. This does not preclude implementing one's own version of a simple technique for convenience in automating large numbers of simulations, or making modifications to existing code where additional functionality is desired. With this in mind, this section reviews several analysis methods and codes that could be used for simulating VADR devices, from a simple total-reflection method that may be implemented easily to highly-sophisticated commercial codes that may not readily be replicated by an individual researcher.

6.2.1 T-Matrix

The transfer matrix (T-matrix) is a straightforward technique that is applicable to the solution of wave equations in the case of normal incidence on a 1-D stack containing one or more interfaces between planar, homogeneous regions. Thus, the technique has found application in electromagnetics, optics, and even quantum mechanics [165]. T-matrices are useful for the analysis of arbitrary stacks of dielectrics because the individual T-matrices for each interface can be simply multiplied together to give the overall T-matrix for the entire system, from which it is straightforward to derive the transmission coefficient.

The general transmission problem for a monochromatic electromagnetic wave impinging on a single 1-D interface is illustrated in Fig. 6.1, where the propagating waves $E(z)$ have the time-independent form of

$$E(z) = E_0 \exp(ikz) \tag{6.1}$$

where z is the position, E_0 the maximum amplitude of the wave, k is the propagation constant in the medium on the left side of the interface ($k = k_1$), or the right side ($k = k_2$).

A T-matrix expresses the waves on the right of a dielectric interface as a function of the waves on the left (see Fig. 6.1a):

$$\begin{pmatrix} C \\ D \end{pmatrix} = T^{(21)} \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} T_{11}^{(21)} & T_{12}^{(21)} \\ T_{21}^{(21)} & T_{22}^{(21)} \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} \tag{6.2}$$

The boundary conditions for a dielectric interface are that the electric (Eq. 6.3) and magnetic (Eq. 6.4) fields must be continuous:

$$E_1 = E_2 \Rightarrow A + B = C + D \tag{6.3}$$

$$H_1 = H_2 \Rightarrow \eta_2(A - B) = \eta_1(C - D) \tag{6.4}$$

where the intrinsic impedances of the regions are $\eta_1 = \sqrt{\mu_1/\varepsilon_1}$ and $\eta_2 = \sqrt{\mu_2/\varepsilon_2}$. Since $\mu_1 = \mu_2$ and $n_1 = \sqrt{\varepsilon_1/\varepsilon_0}, n_2 = \sqrt{\varepsilon_2/\varepsilon_0}$, Eq. 6.4 can be rewritten in terms of refractive indices

$$H_1 = H_2 \Rightarrow n_2(A - B) = n_1(C - D) \tag{6.5}$$

The T-matrix entries are generated from the standard equations for reflection and transmission at a dielectric interface

$$T_{11}^{(21)} = r = \frac{n_2 - n_1}{n_1 + n_2} \tag{6.6}$$

$$T_{22}^{(21)} = -r = \frac{n_1 - n_2}{n_1 + n_2} \tag{6.7}$$

$$T_{12}^{(21)} = t_r = \frac{2n_1}{n_1 + n_2} \tag{6.8}$$

$$T_{21}^{(21)} = t_f = \frac{2n_2}{n_1 + n_2} \tag{6.9}$$

where r is coefficient of reflection , t_f the coefficient of forward transmission, t_r the coefficient of reverse transmission, and n_1, n_2 the refractive indices of the two materials.

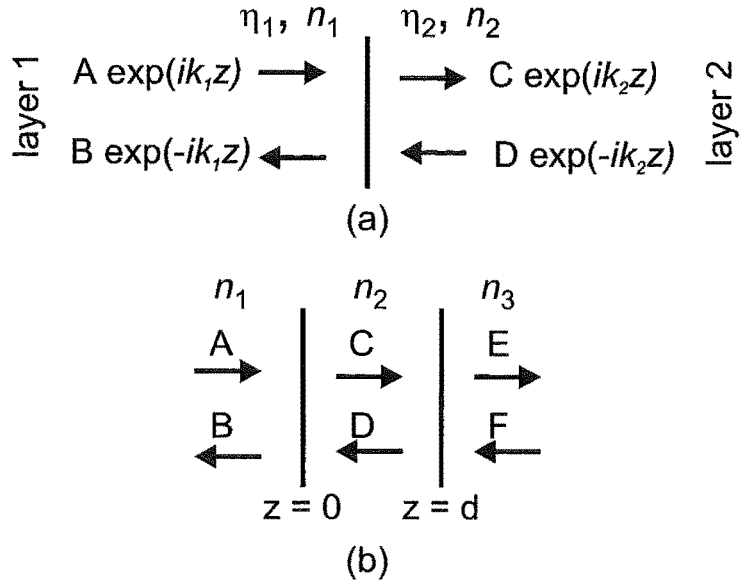


Figure 6.1 The T-matrix describes the waves on the right hand side of a dielectric interface as a function of the waves on the left. (a) single interface between two dielectric layers with refractive indices n_1 and n_2. (b) multiple interfaces.

Thus, the T-matrix for a single interface in terms of refractive indices is:

$$\begin{pmatrix} C \\ D \end{pmatrix} = \frac{1}{n_1 + n_2} \begin{pmatrix} n_2 - n_1 & 2n_1 \\ 2n_2 & n_1 - n_2 \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} \tag{6.10}$$

For a stack with more than one interface, such as the stack in Fig. 6.1(b) with 3 layers, each additional T-matrix must be translated to account for the associated interface's displacement d from the origin (z=0), for example:

$$T^{(32)}(d) = \begin{pmatrix} e^{-ik_3 d} & 0 \\ 0 & e^{ik_3 d} \end{pmatrix} T^{(32)}(0) \begin{pmatrix} e^{ik_2 d} & 0 \\ 0 & e^{-ik_2 d} \end{pmatrix} \tag{6.11}$$

The overall T-matrix for the stack can then be calculated by multiplying together the T-matrices for each of the interfaces:

$$T^{(31)} = T^{(32)}(d) T^{(21)}(0) \tag{6.12}$$

The reflection r and transmission t coefficients are recovered from the overall T-matrix of the stack by setting the incident wave on the left hand side to unity and the incident

wave on the right hand side to zero:

$$\begin{pmatrix} t \\ 0 \end{pmatrix} = T^{(31)} \begin{pmatrix} 1 \\ r \end{pmatrix} \tag{6.13}$$

which gives r and t as:

$$r = \frac{-T_{21}}{T_{22}}, \qquad t = \frac{T_{11}T_{22} - T_{12}T_{21}}{T_{22}} \tag{6.14}$$

The T-matrix equations may be implemented in any suitable language or scripting environment, for example Matlab. The computational load is so low that many devices may be analysed in just seconds. For an example of the results provided by the T-matrix, see the reflection coefficient of the 1-D multilayer stack of GaAs and air, presented in Fig. 2.22(a). The T-matrix technique is used further in Chapter 7, where its application to the VADR structure is explained.

6.2.2 Finite Difference Time Domain Methods

The Finite Difference Time Domain (FDTD) method is used extensively for solving electromagnetic problems [166]. Two features make it particularly attractive. Firstly, the technique is simple. Secondly, since the method tracks the time-varying fields in a volume of space, the FDTD results are easily visualised, and may provide physical insights. FDTD solves Maxwell's time dependant curl equations directly,

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial \mathbf{t}} \tag{6.15}$$

$$\nabla \times \mathbf{H} = \sigma \mathbf{E} + \epsilon \frac{\partial \mathbf{E}}{\partial \mathbf{t}}. \tag{6.16}$$

The region to be analysed must be finite, and is discretised into two interleaving grids. One grid contains the points where the electric field is calculated, the other the magnetic field. The grids are offset by a half grid in all dimensions. The basic implementation requires a cubic grid, but there have been advances in the use of rectilinear (varying-size grids) to better model structures with fine detail on large structures, and conformal grids to better model curved surfaces. A first order central difference

technique is used, yielding second order accuracy. For example,

$$\begin{aligned}\frac{B_x^{(n+\frac{1}{2})}(i,j+\frac{1}{2},k+\frac{1}{2})-B_x^{(n-\frac{1}{2})}(i,j+\frac{1}{2},k+\frac{1}{2})}{\Delta t}\\ = \frac{E_y^n(i,j+\frac{1}{2},k+1)-E_y^n(i,j+\frac{1}{2},k)}{\Delta z}\\ -\frac{E_z^n(i,j+1,k+\frac{1}{2})-E_z^n(i,j,k+\frac{1}{2})}{\Delta y}\end{aligned} \tag{6.17}$$

where Δt is time step, and Δy and Δz are the grid dimensions in the y and z dimensions. The electric field components E_y, E_z are known from previous calculations. In the next iteration the electric fields are solved using a similar expression. The time-marching procedure is sometimes known as "leap-frogging".

In the first implementations of FDTD, an initial field was specified throughout the domain, however this form of excitation was non-compact and taxed the limited computer resources available at that time, restricting excitation to a pulse. The calculation was run until the energy of the pulse had dissipated, and the transient fields were analysed to give a frequency response. Later implementations use a zero initial field throughout, with a smaller source field specified somewhere in the domain, that is updated each time step. This is a more compact source, which provides the additional option of giving a continuous wave excitation. The continuous wave calculation is run until the solution converges to a steady state solution for the EM fields. With appropriate post-processing, this yields the reflection and transmission coefficients (in one form or another) for a single frequency. If information is only required for a single frequency, then the continuous wave simulation is preferable as it often requires less computation resource and time to give results.

The size of the grid and the time step directly influence the speed of computation and the accuracy of the results. The grid size must be small enough that it can accurately represent the system being simulated, but more so, small enough that the wavelength is sampled with sufficient resolution throughout the domain. Ordinarily, the cell size for a cubic grid is

$$\Delta_x \leq \lambda_n/10 \tag{6.18}$$

where all sides of the cell are the same length ($\Delta_z = \Delta_y = \Delta_z$), λ_n is the material wavelength in the most optically dense region of the simulation domain ($\lambda_n = \lambda_0/n_{max}$, where λ_0 is equal to the shortest free space wavelength of significant power). The time step is set in proportion to the chosen cell size such that the wave cannot propagate between any two mesh points faster than the equations are updated. Otherwise oscillations are introduced that have a wavelength equal to two grid cells and an exponential growth rate [166]. The exponential growth rate of these instability oscillations ensures

that they soon dominant the simulation and render the results invalid. The related Courant stability factor S is defined as

$$S = \frac{c\Delta t}{\Delta x}, \tag{6.19}$$

and for values of $S > 1$, the time step is too large relative to the cell size and as a result the simulations are numerically unstable.

Thus, the number of calculations required is proportional to

$$N \propto \frac{X.Y.Z}{\Delta_x^3} \times \frac{cT}{S\delta_x}, \tag{6.20}$$

where X, Y, Z are the overall simulation domain dimensions, T is the overall time, and S is the Courant stability factor. Therefore the overall computation time is increased by a factor of eight if the cell size is halved. As a consequence, a compromise must often be made between accuracy and computation time. With a current PC, trivial simulations of 1-dimensional structures take several seconds whereas 3-dimensional simulations of structures that are dimensioned on the order of several wavelengths can take weeks.

In its most basic form the FDTD method makes few assumptions, and therefore has the potential to produce accurate results. The trade-off in making few electromagnetic assumptions is a high computational load.

6.2.3 Method of Moments

The "method of moments" (MoM) is a process for turning linear equations with continuous variables into matrix equations that may be solved with the aid of a computer. As such, it is well-suited to electromagnetic problems [167]. There are a variety of names for methods that are encompassed by this process, for example "Galerkin methods", "Residual Methods" and "Boundary element Method".

In MoM, the region in which the electromagnetic fields are to be solved must be discretised in space as for the FDTD technique. However, the fields themselves are also discretised (expanded), according to a Fourier or other basis function, over the whole spatial domain. The details of spatial discretisation and field expansions are specific to each implementation, but in general, an inhomogeneous equation is solved approximately, so that only a finite number of terms are required in the field expansion. For electromagnetics, the solution begins with an inhomogeneous equation is often of the form

$$L(f) = g \tag{6.21}$$

where L is a linear operator (in electromagnetic MoM problems L is usually an integral operator), g is a known excitation or source, and f is the unknown response.

Once the system has been reduced to a matrix equation, a matrix inversion process is performed by computer. Since matrix inversion is computationally-costly (order $N^{log_2 7}$ with efficient methods such as the recursive Strassen algorithm [168]), there is a trade-off between computation time and solution accuracy (number of terms in field expansions). Advantageously, however, the system is represented for an arbitrary excitation, so it is conveniently solved for the responses at all frequencies of interest at once [169].

Unfortunately, the only MoM implementation available to the author was not suitable for simulating VADR. AntennaSolver is freely available [170], and based on the well-known "Numerical Electromagnetic Code" (NEC2) code developed by the Lawrence Livermore Laboratory, California [171]. Since it is designed for modelling of antennas, geometries are constructed of metals and it is not possible to specify regions of arbitrary dielectric materials such as those required for modelling VADR. It should be noted that this is not the only the implementation of MoM in existence – but no others are reviewed since MoM was not used in the present study.

6.2.4 Finite Element Methods

The Finite Element Method (FEM) is an integral method, and shares some methodology with the Method of Moments. Like FDTD and MoM, a domain is specified that contains the problem space, however the domain is not gridded, but rather meshed with series of tetrahedral elements formed from four equilateral triangles. The dimensions of these mesh elements may be freely varied within the domain so that finer structures may be adequately represented without unnecessarily increasing the number of cells in sub-volumes with larger-scale features, and often this process is performed iteratively by the FEM solver during a simulation until convergence is reached. For electromagnetic problems in 3-D, an energy functional is minimised to solve the following wave equation [172]

$$\nabla \times \left(\frac{1}{\mu_r} \nabla \times \mathbf{E}(x,y,z) \right) - k_0^2 \varepsilon_r \mathbf{E}(x,y,z) = 0 \tag{6.22}$$

where $\mathbf{E}(x,y,z)$ is the complex electric field within the simulation domain, μ_r is the position-dependent complex relative permeability and ε_r is the position-dependent complex relative permittivity and k_0 is the free-space constant $\omega\sqrt{\mu_0 \varepsilon_0}$ where ω is the angular frequency $2\pi f$.

Initially, FEM solvers were restricted to solving solutions within a closed geometry. This would not be appropriate for the VADR device because it operates in free-space. Recently, implementations of FEM such as Ansoft's HFSS have added radiation boundary conditions that are similar in principle to their FDTD counterparts [173]. Unfortunately this particular implementation of FEM was the only one potentially available to

the author for this work, and it did not implement periodic boundary conditions. This unacceptably increased the size of the simulation domain because it would have been necessary to include a significant number of VADR grooves in the simulation domain in order to obtain an accurate result. Since the entire FEM simulation domain is meshed, this prohibitively increased the length of the computation.

6.2.5 Generalised Multi-pole Technique

The Generalised Multi-pole Technique (GMT) is another integral technique, and operates in the frequency domain [174]. However, it does not suffer the same simulation domain size-limitations of the other integral methods discussed here (MoM, FEM) because it does not discretise the entire simulation domain but rather just the boundary conditions. This is an advantage for electrically large structures with simple geometries, especially those that require large regions of free space in the simulation domain such as VADR, since it results in shorter simulation run-times.

The boundary conditions, including dielectric interfaces, are discretised by placing field expansions nearby and specifying points on the boundary where the field expansions must match. In the GMT, these field expansions are known analytic solutions of the electromagnetic field equations, and are often based on sine, cosine, Bessel, Hankel or Neumann functions, depending on the geometry of the boundary. The most commonly used expansion is the multipole, hence the name of the technique.

The field expansions build up a representation of the fields in the different areas of the simulation domain by acting as fictitious sources for the EM fields. For time-harmonic problems, the coefficients of the field expansions are calculated by solving the Helmholtz equation

$$\nabla^2 \mathbf{U} + k^2 \mathbf{U} = 0, \tag{6.23}$$

where $\mathbf{U}$ is the electric or magnetic field of interest, and k the wavenumber. Once the coefficients are determined, the field can be calculated at any point.

An advantage of GMT is that as only the boundaries are discretised, the computational load for a given problem can be substantially lower than for FDTD or FEM where the entire simulation domain must be discretised. The disadvantage of GMT is that it is sensitive to the placement of the expansions, and no reliable algorithm exists to automate this procedure. Time and experience are required to develop a 'feel' for the process, particularly if accurate results are to obtained within a reasonable computation time. The computation time is directly related to the number and order of expansions, which reflects how finely the border has been discretised. Various implementations of the GMT restrict themselves to certain combinations of the expansions, and a multiple multipole (MMP) incarnation of GMT called MaX-1 was in fact used

for the original numerical simulations of the VADR device [65].

6.2.5.1 Multiple Multipole

The available MaX-1 multiple multipole (MMP) implementation of GMT was developed by Christian Hafner at the Swiss Federal Institute of Technology in Zurich [175]. This code, named MaX-1, solves within a 2-D problem space and assumes that the structure has continuous translation symmetry in the third direction. This is suitable for the modelling of the VADR device, and the MMP solver in MaX-1 was in fact used for the MMP calculations presented in the VADR device proposal [65]. Results of this work have already been reproduced in Fig. 2.19.

It is worth noting that assumptions made in the original VADR proposal led to a minimum number of boundary conditions being included in the simulation model. The work required in this chapter would require the number of boundary conditions to be increased, which would non-trivially complicate the manual placement of the field expansions. This is no reason to avoid the technique, especially if other aspects such as the fast computation time were advantageous. For example, periodic outer-boundary conditions are available with MaX-1, but not with the Ansoft HFSS implementation of FEM. However, MaX-1's limitation to solving in 2-D would prevent future investigations into improved VADR devices from exploring devices with orthogonal grating vectors (such devices will be discussed in Chapter 7) since this would violate the requirement for continuous translation symmetry in the *third* dimension. Even if a 3-D solver were available, manual optimisation of the field expansion placement within a 3-D space would impractical for most non-trivial problem geometries.

6.2.6 Rigorous Coupled Wave Analysis

The rigorously coupled wave analysis (RCWA) formulated by Moharam and Gaylord [176] is a differential method, and unlike the Max-1 MMP implementation, it is not limited to just 2D structures. However, it does still suffer some limitations in this regard, as will be discussed later in this section. Despite this, it is an attractive technique to use because it is relatively straightforward, noniterative (unlike some FEM implementations, for example), and deterministic.

Moharam and Gaylord initially developed RCWA for holographic gratings, but they and other researchers have since extended it to model surface relief gratings such as the VADR V-grooves. The RCWA algorithm calculates the diffraction efficiencies of a stack of binary layers interposed between two semi-infinite regions of either substrate or air. Even continuous grating profiles may be constructed from a number of binary layers, as shown in Fig. 6.2; the accuracy of the calculation improves as more layers are used in the stepped-staircase approximation to the profile, although this has the disadvantage of also increasing the computation time. When the grating geometry is

modelled exactly, the accuracy is only limited by the number of diffracted orders for which the diffraction efficiencies are calculated.

Since energy must be conserved, greater accuracy is achieved with more, rather than less, orders retained in the calculation. Each propagating order should be retained in the calculation, plus some evanescent orders. For a dielectric grating, only a few evanescent orders needs to be retained since the higher order evanescent field magnitudes decay rapidly.

The RCWA algorithm proceeds as follows. First, the permittivity profile within each layer is expanded in a Fourier series, and then fields within each of the layers are calculated with full-vector Maxwell's equations. Then, the electromagnetic boundary conditions, which are that tangential electric- and magnetic-field components must be continuous, are applied sequentially in the output region, between the individual grating layers and finally in the input region. This yields the desired complex diffraction efficiencies. Since this calculation necessarily includes the zeroth diffracted order, the technique may also be applied to solving geometries with subwavelength gratings.

The ability to model subwavelength gratings is particularly useful when the structure is one that is not accurately solvable by homogenisation with effective medium theory (EMT), and has the advantage of being a fast calculation since only a small number of orders must be retained. Additionally, the Fourier expansion of the permittivity implies a periodic grating is being analysed, and this provides an efficient solution for devices like VADR with periodic structures.

There has been some discussion in the literature with regards to obtaining a numerically stable algorithm for deep gratings, for metallic structures and for gratings consisting of more than one grating layer. Neither of the first two situations arises in the VADR analysis since the structure is dielectric and has shallow (sub-wavelength depth) gratings. However, the last condition is relevant, and it appears that there is a trade-off between computation speed and risk of experiencing numerical instabilities, particularly for lossy materials [176]. Therefore, it might be expected to take longer to calculate a lossy grating.

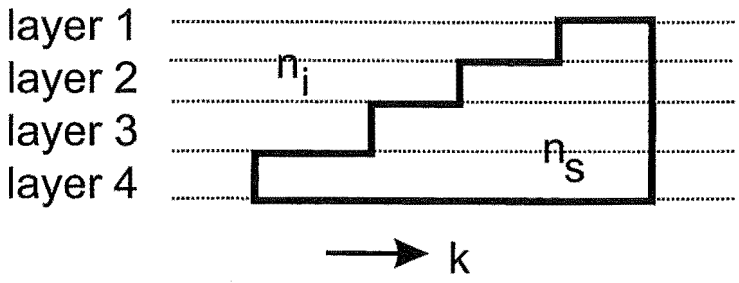


Figure 6.2 An arbitrary grating profile (here, a sawtooth) that has been decomposed into binary layers for use with the RCWA technique.

As indicated at the start of this section, there are some limitations in regard to structures with crossed gratings (orthogonal grating vectors). Ordinarily, the computation time grows with N^3 where N is the number of orders retained in the calculation. However, for crossed gratings, the computation time grows with order N^6, and therefore it is only practical to solve for a few orders [177]. This limitation may be eased slightly with the aid of a high performance computer.

At least two commercial RCWA programs are commercially available, including Fluckiger's "Gsolver" [178] and Goray's "PC Grate" [179]. Gsolver was available for use by the author, and is described in following section.

6.3 GSOLVER PROGRAM

The Gsolver program implements the RCWA technique, and was developed by David Fluckiger of the Grating Solver Development Corporation [178]. Gsolver is made available under a commercial license. Gsolver also solves for cases with crossed gratings (orthogonal grating vectors) and this is useful for one of the improved version of VADR.

The user interface, as shown in Fig. 6.3, is straightforward, and split into three windows. In the first window ("Editor"), details of the geometry are entered. This includes the wavelength, periodicity and profile of the permittivity within each of the layers in the geometry. In second window, "Run", details of the number of orders to be retained in the simulation are set, with the program analysing the structure and suggesting the minimum. Also, advanced features may also be specified such as calculations across a range of incident wavelengths, or other input parameter. The third window, "Results", presents the calculated transmission and reflection coefficients for each order in the calculation. More detailed results are obtained by specifying that the results should be written to file during the simulation.

6.4 VALIDATING RCWA AGAINST EMT AND FDTD

Even though it is expected that RCWA would accurately model subwavelength features, it is important to confirm this by detailed comparison against calculations made with effective medium theory (EMT). This also has the advantage of allowing the regions of validity of EMT to be explored. The effects investigated here are related in particular to the period-wavelength and depth-wavelength ratios of a simple subwavelength grating. This gives confidence in the use of EMT within its useful range, but perhaps more importantly for the present work, gives confidence in the use of the rigorous techniques in the analysis of subwavelength gratings.

The 1-dimensional lamellar subwavelength grating analysed in this section is shown in Figure 6.4, and had the following fixed properties: fill factor $\xi = 0.5$, $n_{gr} = 1$ (air) and $n_{rd} = 3.42$ (silicon). The incident free-space wavelength was $\lambda = 3$ mm ($f = 100$ GHz).

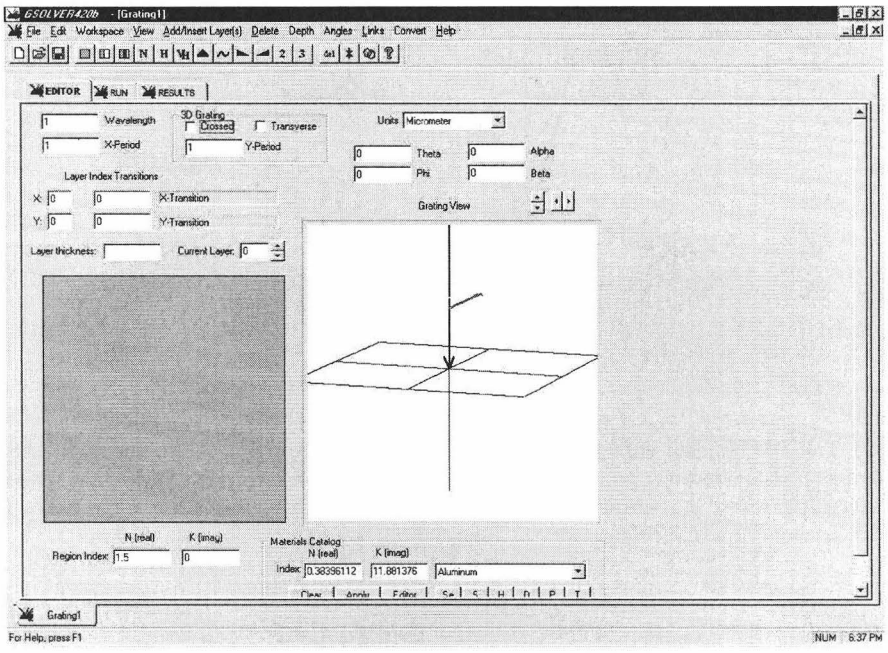


Figure 6.3 The Gsolver user interface allows the geometry to be interactively edited.

Throughout this section, the medium above and below the grating is assumed to be air with $n = n_{gr} = 1$.

In the first analysis, the depth was kept constant at $h = 100$ μm, while the period Λ was varied from 1μm to 500μm. These values represent the smallest and largest values for which rectangular-grooved VADR devices might reasonably be expected to be analysed with EMT. A small period such as $\Lambda = 1\mu$m approaches the long wavelength limit, while $\Lambda > 438$ μm ($\lambda_0/2.n_{rd} = 438$ μm) can support non-zero propagating orders in the substrate (which, if present, reduce the accuracy of EMT). The $\Lambda = 500$ μm period of the VADR devices was designed to reduce the aspect ratio of the grooves to suit the fabrication procedures.

The grating was simulated using RCWA and EMT/T-matrix analysis for zeroth order, second order and exact Rytov EMT approximations. The magnitude of the electric field transmission coefficient is plotted against the period in Fig. 6.5, for zeroth, second order and Rytov EMT, and Gsolver RCWA, with TE in Figure 6.5(a) and TM in Figure 6.5(b). As expected, all four techniques were in agreement in both polarisations at small periods, also known as the long-wavelength limit. The zeroth order EMT neglects the period, so the corresponding transmission coefficient was constant at 0.88195 for TE and 0.99623 for TM for all periods. Second order EMT, and the more accurate Rytov EMT, both take the period into account, and the effect of this is clear in both polarisations.

For the TE polarisation, the rigorously calculated transmission coefficient, from RCWA, agreed well with the second order and exact Rytov EMT, to within 0.006%,

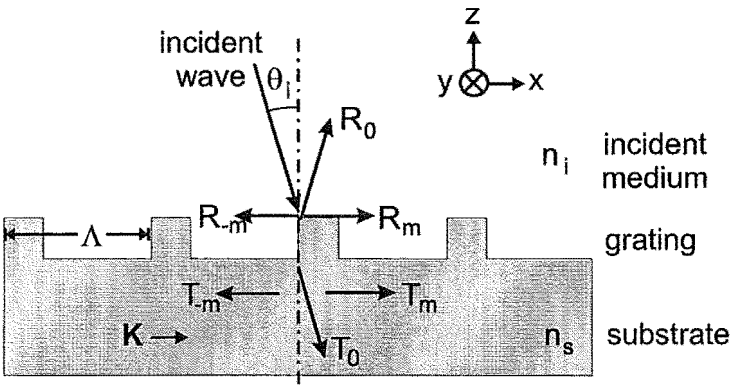


Figure 6.4 The 1-D lamellar subwavelength grating problem. A grating of period Λ with a rectangular profile has rods of depth h and width of $f\Lambda$. The rods have refractive index n_{rd}, while the air gaps have refractive index n_{gr}. Above the substrate the refractive index is n_1, while below it is n_2. The illumination is normally incident.

for periods of 1 μm up to approximately 100 μm. At larger periods, the predictions diverged. At a period of 500 μm, the relative error is 1.3% between the exact Rytov EMT and the RCWA transmission coefficients.

For TM, the slope of the second order and Rytov EMT results was not consistent with the slope of the RCWA results, even at small values, which was unexpected. The relative error between the transmission coefficients predicted by Rytov EMT (0.99377) and RCWA (0.98760) at a period of 496μm was 0.6%.

Thus, it may be seen that there is a divergence even at very small periods, where none is expected. This is a concern, and the cause needs to be identified.

The second order and Rytov EMT assume that the grating is sufficiently deep that any evanescent fields have decayed away before reaching the interface at the bottom of the grating, and thus only the propagating fields are considered. For shallow gratings, especially those less than a quarter of the free space wavelength deep, the evanescent fields may not have fully decayed, and the grating properties can vary strongly as a result. For accurate analysis of shallow gratings, those less than a quarter wavelength deep, the evanescent fields must be accounted for, and the Lalanne EMT provides a formulation for this [90]. For accurate modelling of the phase response in extremely shallow gratings, those less than a tenth of a wavelength deep, an effective thickness must be calculated in addition to the usual effective permittivity [180]. It appears from these references that the figures of one quarter and one tenth of a wavelength are best treated as broad guidelines rather than firm boundaries for the use of one EMT formulation or another.

To examine the effect of depth on the effective permittivity, the same grating was analysed in RCWA with a fixed period, $\Lambda = 500\mu$m, with grating depths of $h = 10 -$

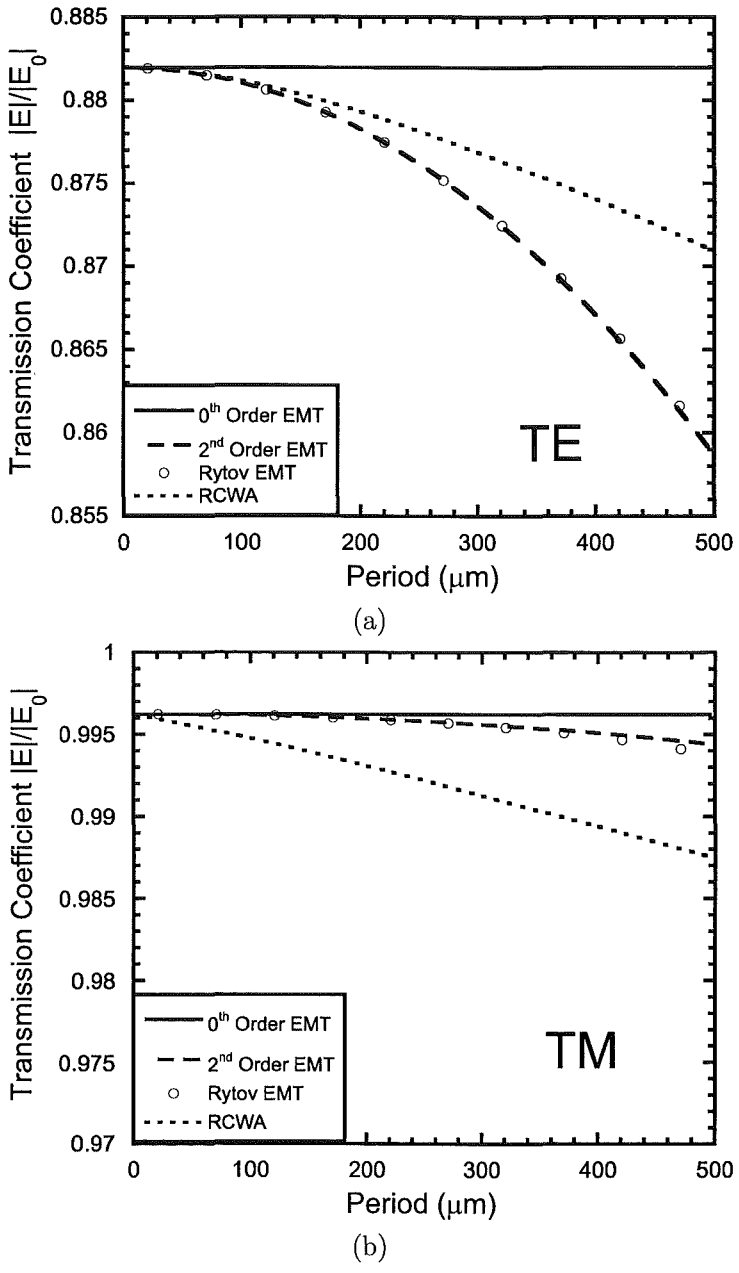


Figure 6.5 The magnitude of the electric field transmission coefficients for a grating of the type in Figure 6.4, where depth $h = 100\mu$m, fill factor $f = 0.5$, $n_{rd} = 3.42$, n_{gr}=1.00, and period Λ varies from 1-500mum. (a) TE polarisation, (b) TM polarisation.

100 μm, in steps of 10 μm. RCWA does not give an effective permittivity directly, however it can be calculated by modelling an equivalent-thickness homogeneous film. By calculating the transmission coefficient of the thin film with permittivities in the interval $n_{gr}^2 = 1$ to $n_{rd}^2 = 11.7$, in steps of 1×10^{-4}, the effective permittivity of the homogeneous film that gives that same transmission as the grating can be identified. The procedure is illustrated by example as shown in Fig. 6.6, although in practice it was automated in software. The transmission coefficients for the homogeneous 50μm layer, with varying permittivity, are plotted in Figure 6.6 as a solid curve, while the transmission coefficient of the actual grating, is plotted as a dashed line. Note that only the permittivity of the homogeneous layer was varied, and that the line representing the transmission of the actual grating is only plotted to aid the illustration of the procedure. The permittivity of the thin film at intersection of the two traces is the effective permittivity of the grating, in this case 6.5283.

The permittivities of the grating at depths 10-100μm, determined using the procedure just explained, are plotted in Figure 6.7, along with the values calculated by zeroth, second order and Rytov EMT. The RCWA values clearly show a variation with the depth of the grating. Note that all other parameters, including the period, were fixed for this part of the analysis. As expected, the zeroth, second order and Rytov EMT predictions fail to vary with the depth and do not accurately represent the grating's effective permittivity. It appears as if the zeroth order EMT and the Gsolver traces would converge in the limit of zero grating depth. However, the zero-depth limit cannot be simulated directly with the RCWA technique since it is treated as a non-existent layer.

Having established that the zeroth, second order and Rytov EMT fail to accurately model the grating's effective permittivity at small depths, and that RCWA predicts that the effective permittivity does vary with the depth, it is clear that there is a need for an EMT formulation that accounts for the grating depth, such as that provided by Lalanne [90] and reviewed in section 2.4.6. Unfortunately, the calculation of an effective permittivity for a lamellar grating in the TM polarisation is not possible with this formulation, since the number of terms in the Fourier expansion of the permittivity is included in the numerator of the TM equation (Eq. 2.27). For discontinuous profiles, an infinite number of Fourier terms are required, and the value of the effective permittivity incorrectly tends to infinity. Thus the remainder of the discussion concerning the lamellar grating, and pertaining to the Lalanne EMT formulation, is necessarily restricted to the TE polarisation.

For the TE polarisation, the effective permittivity predicted by the Lalanne formulation is plotted in Figure 6.8 as open squares, along with the traces from Figure 6.7 for comparison. The agreement is exact at a depth of 10μm, with the error slowly increasing to 0.04, or 0.6%, which compares favourably to errors of 0.35 - 0.70 (5.2% - 10.5%) for Rytov EMT.

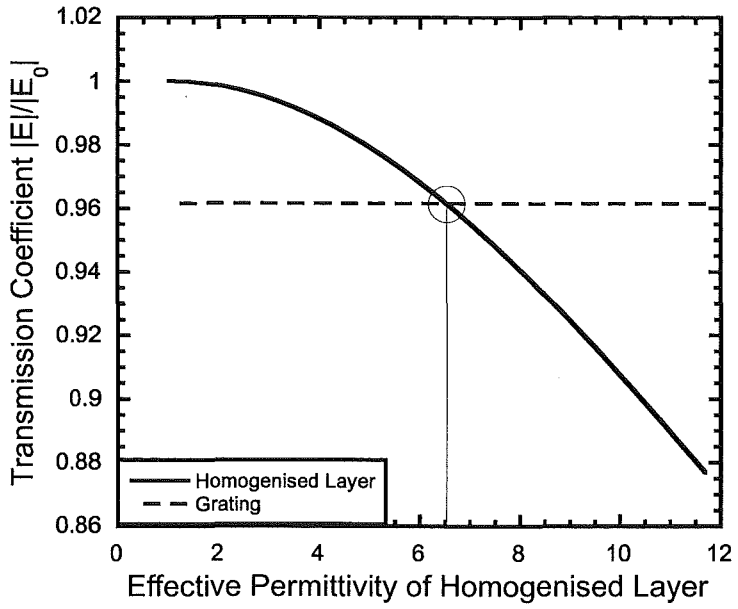


Figure 6.6 The effective permittivity of the grating can be found by comparing its transmission (calculated by RCWA) with the transmission coefficients of a homogeneous layer of the same depth, modelled with permittivities in the range $n_g^2 r$=1 to $n_r^2 d$=11.6964. The effective permittivity of the grating is read from the graph where the two curves meet. Note that only the permittivity of the homogeneous layer is being altered, and not that of the grating itself.

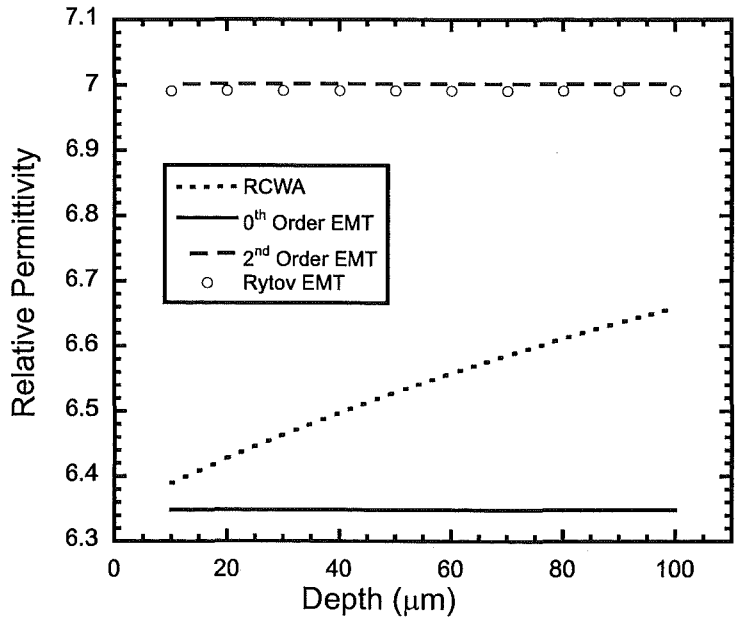


Figure 6.7 The effective permittivity of the grating as a function of depth $h = 10 - 100\mu$m. RCWA clearly shows that the effective permittivity is depth dependant at these small grating depths, whereas the permittivity predicted by zeroth, second order and Rytov EMT do not show the depth dependence.

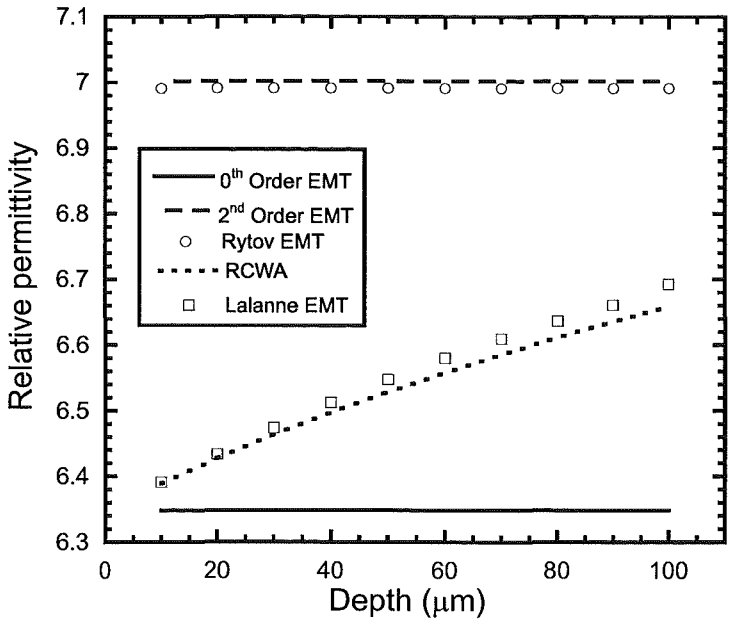


Figure 6.8 The effective permittivity of the grating as a function of depth $h = 10 - 100\mu$m, with Lalanne EMT values plotted as open squares, and showing good agreement with values determined from RCWA.

To show that the Lalanne EMT formulation also deals with variation in the period accurately, the technique was used to analyse the fixed depth (100μm), variable period ($1 < \Lambda < 500\mu$m) grating that was first analysed in this section (with the results in Figure 6.5). The transmission coefficients calculated using Lalanne EMT are plotted in Figure 6.9 as open squares, and show a great improvement over the transmission coefficient predicted by Rytov EMT. The maximum error between RCWA and Lalanne EMT was 0.001 (0.1%) at a period of 500μm, and was approximately a tenth of the error for Rytov EMT, 0.011 (1.2%), also at a period of 500μm.

As a final confirmation, another rigorous technique, FDTD, is used to corroborate both the Lalanne EMT and RCWA results for the TE polarisation, and the RCWA results for TM. Again, the fixed depth grating was analysed, so that the results could be compared to Figure 6.9 for TE, and Figure 6.5(b) for TM. A 2-dimensional FDTD domain was configured as shown in Figure 6.10, of size 20 cells in x by 1200 cells in z for grating periods 200 - 500μm, and 20 by 2200 cells for grating periods 10 - 100μm. No gratings were simulated with periods between 100 - 200μm. The change in domain size was to ensure that the grating was at least a half wavelength from each of the absorbing Perfectly Matched Layer (PML) boundaries, as the behaviour of these boundaries is not reliable in the presence of evanescent fields.

The grating was periodic in x, with one period of the grating present in domain. An electric field was launched at the top of the domain, and propagated in the $-z$

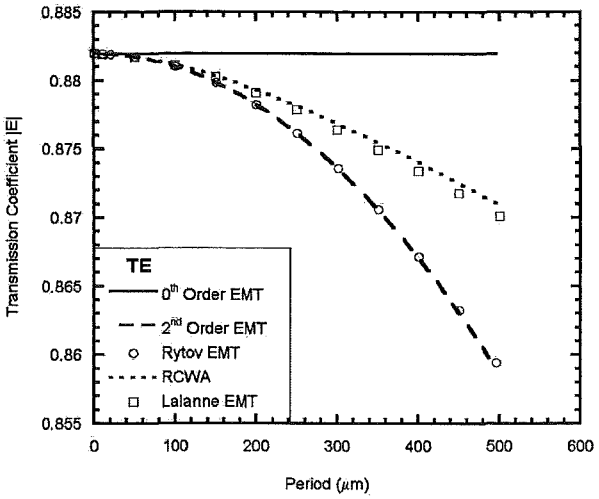


Figure 6.9 The Lalanne EMT (open squares) predicts the transmission coefficient of a 100μm deep grating, with period $1 < \Lambda < 500\mu$m, with less than a tenth of the error of Rytov EMT, as compared to RCWA.

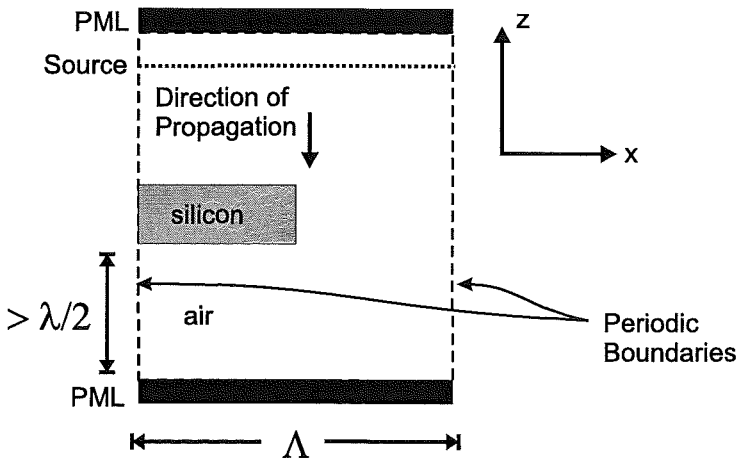


Figure 6.10 The FDTD simulation domain was one period wide, and periodic in x. The plane wave excited by the source propagates in the negative z direction. PML absorbing conditions were used. The TE and TM directions are shown.

direction. The arrangement of the Yee cell in the Tempest FDTD code presented the silicon rectangle of the grating to the TE polarisation with one less cell in the x-direction than for the TM polarisation, giving a fill-factor of 0.45 rather than the required 0.5. In this analysis, separate models were used for the TE and TM polarisations. The error could be reduced by increasing the number of cells in the x direction, at the expense of increased computational time. The FDTD and Lalanne EMT transmission coefficients are plotted in Figure 6.11, along with the RCWA, zeroth, second order and Rytov EMT results.

Unfortunately, it would not be possible to use the Lalanne formulation in place of the exact Rytov EMT in the analysis of a VADR device because it has not been proven for stacked gratings, and the results for the TM polarisation have not been validated. Without results for the TM polarisation it is not possible to determine the retardance of a VADR device.

In summary, discrepancies between the results of RCWA and the second order and exact Rytov EMT analyses of a lamellar grating were investigated, in order to check the validity of the RCWA simulations for subwavelength gratings. An example was shown of the depth dependence of the effective permittivity of a shallow subwavelength grating. It was also shown that Rytov EMT provided an inadequate model in these circumstances. In order to get good agreement between EMT and RCWA it was necessary to use the Lalanne EMT formulation, that accounted for not only the period of the grating, but also for the evanescent fields that were present due to shallow nature of the grating. The Lalanne EMT provided a much improved approximation of the grating's properties such as effective permittivity and transmission coefficient.

A second rigorous technique, FDTD, was used to corroborate the results produced by RCWA, and showed excellent agreement in both TE and TM polarisations. There can be good confidence in the RCWA results for the VADR devices because it has been shown that there is close agreement between the results of three different techniques (RCWA, EMT and FDTD), even in the difficult case (for EMT) of a shallow grating with a large period.

6.5 VADR MODEL

Two main considerations were made in constructing the VADR model, firstly in representing the structure, and secondly representing the material properties. The Gsolver program can only analyse lamellar gratings, thus it is necessary to use a staircase approximation to represent the v-grooves as shown in Figure 6.12. The results are not presented here, but simulations of single VADR plates were performed with varying coarseness to determine the number of steps required to obtain an adequate representation. The results converged for structures with 8 or more layers. Using ten levels yielded a convenient layer thickness to work with in model creation. This was useful

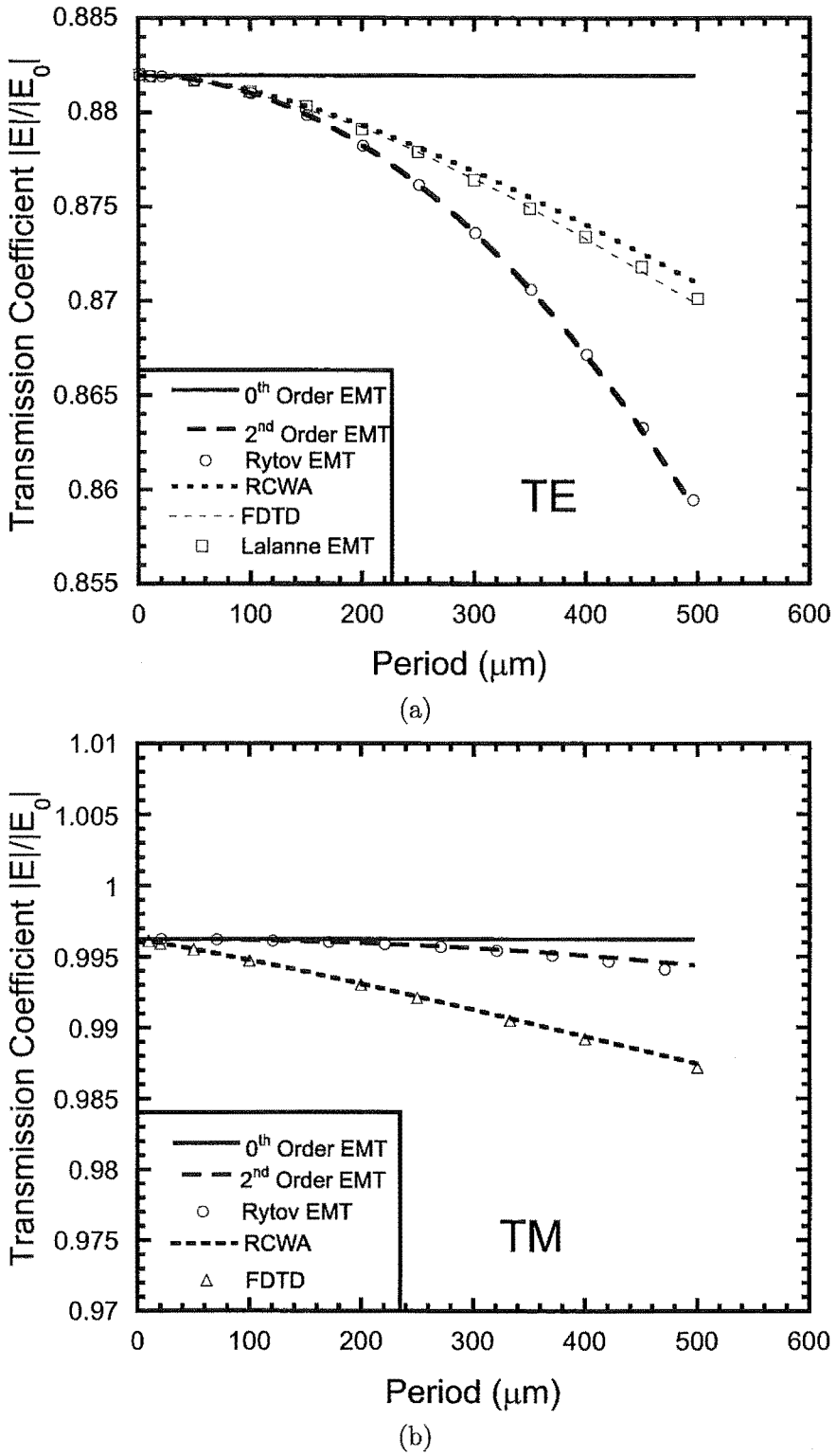


Figure 6.11 The magnitude of the electric field transmission coefficients for a grating of the type in Figure 6.4, where depth $h = 100\mu$m, fill factor $f = 0.5$, $n_{rd} = 3.42$, n_{gr}=1.00, and period Λ varies from 1-500mum. The two rigorous methods (FDTD and RCWA) agree with the Lalanne EMT formulation that includes the grating depth in the calculation of the effective permittivity.

because a separate model was required for each step in the separation within the interlock region, and these were created by hand in a text file description of the geometry which was subsequently imported into the program. The text file format was specific to Gsolver, and was the original geometry input method for the early versions of the code. Using more layers provided little gain in accuracy but drastically lengthened the process of building the models, as the program deals with a stack of layers and has no concept of the two wafers being interlocked. The simulations modelled the interlock region (0 – 350μm) in 35μm steps, to match the step height in the staircase approximation of the grooves so that model construction was straightforward. The remainder (350 – 3000μm) was modelled the device in 50μm steps.

All that remains is to select a value for the refractive index of the silicon, taking into account the lossy nature of the 10 – 20 Ω-cm silicon used in the experiment.

6.5.1 Silicon refractive index

It is difficult to predict the loss of a silicon wafer directly from its specified resistivity, so it is the usual practice to measure the loss experimentally. Unfortunately, no measurements were found in the literature for the 10 – 20 Ω-cm p-doped silicon used to fabricate this device. The nearest values found were for 9.0 Ω-cm [181] and 50 – 100 Ω-cm silicon [182], as summarised in Table 6.1. The complex permittivity for the 50 – 100 Ω-cm silicon was a tabulated value, but the value for the 9.0 Ω-cm was determined from the simplest Drude model with fitted parameters.

The Drude model treats free carriers in a solid as classical point charges that are subject to random collisions. For the particular model used here to determine the complex permittivity of 9.0 Ω-cm silicon, it was assumed that the collision damping of the carriers was independent of their energy. Thus, the model was

$$\varepsilon = \varepsilon_\infty - \frac{\omega_p^2}{\omega(\omega + i\Gamma)} \tag{6.24}$$

where ε is the complex permittivity at angular frequency ω, ε_∞ is the contribution from the dielectric which may be obtained from measurements of undoped silicon to be 11.66 for 0.1 - 2.0 THz and $\Gamma = 1/\tau$ is the collision damping rate, where τ is the average collision time; ω_p is the plasma angular frequency defined by $\omega_p^2 = Ne^2/\varepsilon_0 m$

Table 6.1 Literature values for the complex permittivity and refractive index of p-doped silicon at 100 GHz

Resistivity Ω-cm	ε'	ε''	n_r	n_i
9.0 [181]	11.50	1.92	3.403	0.2825
50 – 100 [182]	11.80 ±0.08	0.25 ±0.02	3.435	0.0364

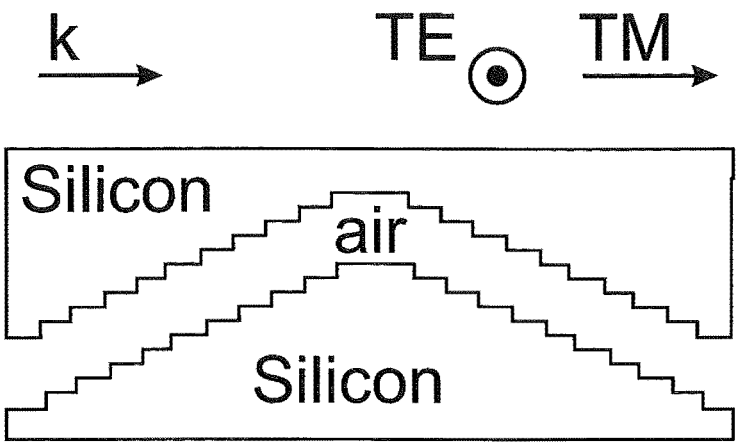


Figure 6.12 The 10-layer staircase model for RCWA simulations of VADR. One period is shown.

where N is the number density of carriers, e is the electronic charge, ε_0 is the free-space permittivity and m is the effective carrier mass [181]. For the 9.0 Ω-cm silicon, the tabulated values were $\omega_p/2\pi = 0.48$ THz, $\Gamma/2\pi = 1.19$ THz, and at 100 GHz, $\omega/2\pi = 0.1$ THz. These values may be substituted in Eq. 6.24 to arrive at the values listed in Table 6.1.

Since the thickness of the plain silicon plates is known to be 550 μm, the etalon may be simulated with the aid of the T-matrix method, RCWA or some other appropriate method. T-matrix simulation results for the transmission magnitude of the etalon are presented in Fig. 6.13. The solid line is the measured result for comparison, whilst the dashed line is the simulated result for the etalon with silicon plates of resistivity 50 – 100 Ω-cm and the dotted line is the simulated result for the etalon with silicon plates of resistivity 9.0 Ω-cm.

As expected, the measured magnitude lies between the two simulated magnitudes, and is is closest to the simulation with the lower resistivity silicon. This indicates that the resistivity of the actual silicon is closer to 9.0 Ω-cm than 50 Ω-cm. Since the measured transmission magnitude is bounded by the two simulated transmission magnitudes, it is reasonable to expect that the loss could be used as a fitting parameter, in order to obtain a more accurate estimate of the silicon.

A series of etalons were simulated with all complex permittivities in the range $11.000 < \varepsilon' < 12.000$ and $0.000 < \varepsilon'' < 2.500$ with a granularity of 0.001. The fit was judged on the basis of the least mean-squared error, giving the fitted value of the complex permittivity to be ε =11.560 + 1.749i. This is equivalent to a refractive index of $n = 3.41 + 0.257$i. The degree of fit obtained may be seen in the plots of the measured and simulated log magnitude of the transmission coefficient in Fig. 6.14. The simulated and measured magnitudes agree 0.5 dB. Perfect agreement is not obtained

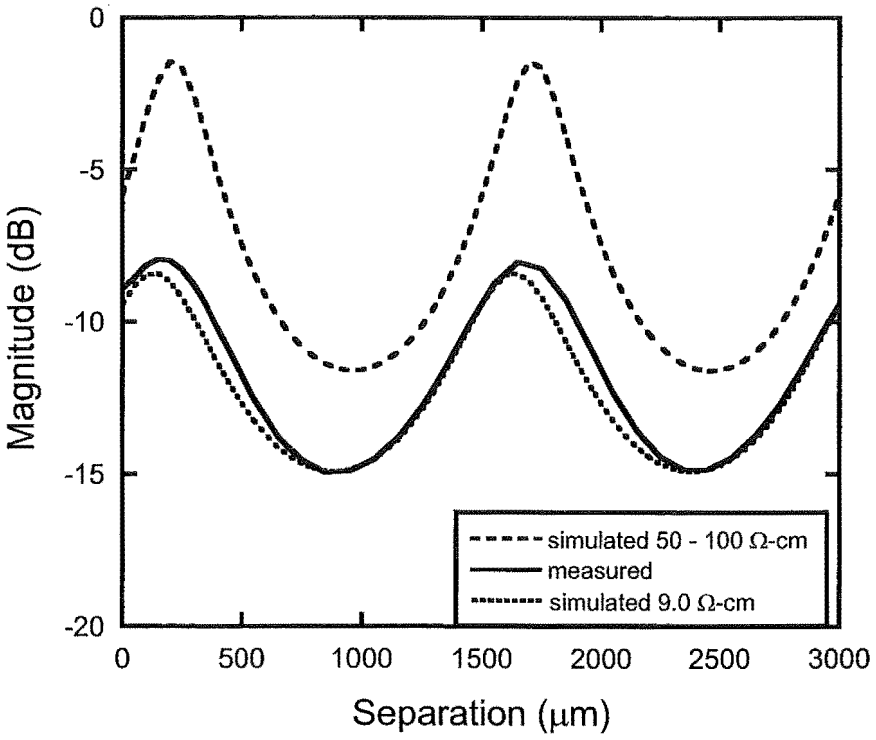


Figure 6.13 Log magnitude of the simulated transmission coefficient at 100 GHz for the plain silicon etalon using tabulated values for the complex permittivity of the silicon substrate.

due to experimental errors such as slight misalignments of the plates. However, the value obtained is expected to give a better fit in the VADR simulations than the tabulated values for 9.0 Ω-cm and 50 Ω-cm silicon.

6.6 PHASE REFERENCE

Since the size of the Gsolver simulation domain is set by the overall thickness of the VADR device, increasing the separation distance causes the simulation domain to increase in size by an equivalent amount (but only in the direction of propagation). The consequence of this is that the phase of the complex transmission coefficient increases with the change in separation by an additional amount equivalent to

$$\Delta\phi = e^{jk(\Delta s)} \tag{6.25}$$

where $k = 2\pi/\lambda_0$ is the propagation constant and λ_0 is the wavelength in free-space. At 100 GHz, $\lambda_0 = 3$ mm, and so a change in the separation distance of 0 – 3 mm is equivalent to a 2π shift of the phase reference point in the simulation domain. This artifact is not present in the measurements, because it is akin to moving the horn antennae further apart as the separation increases. Therefore, to compare the simulation results directly with the measurements, at 100 GHz, the phases (in degrees) of the simulated

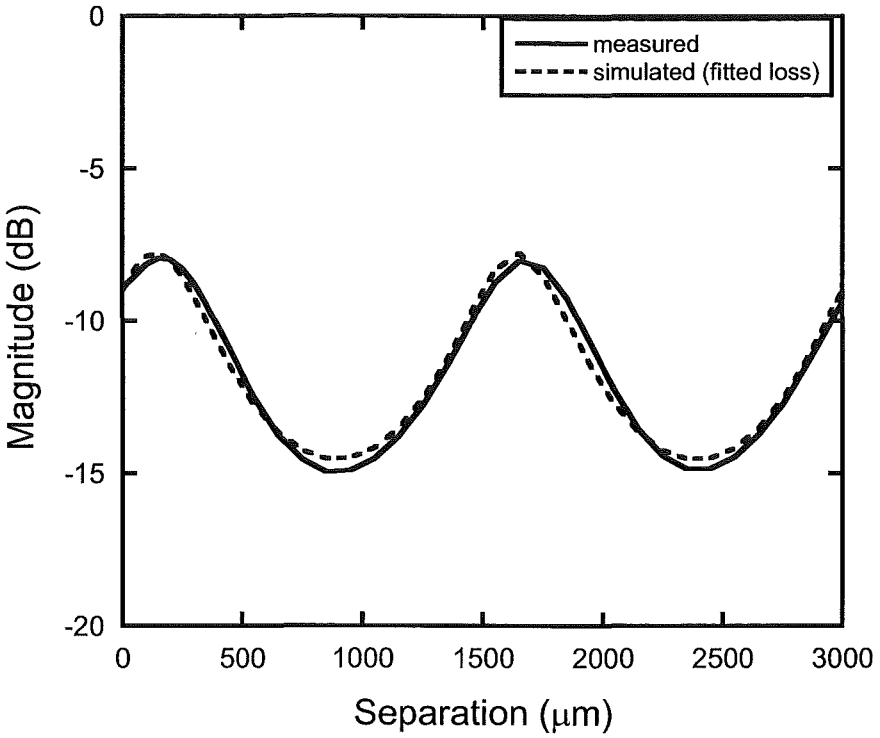


Figure 6.14 Log magnitude of the simulated transmission coefficient at 100 GHz for the plain silicon etalon using a fitted value for the complex permittivity of the silicon substrate.

complex transmission coefficients are adjusted according to the following formula

$$\phi_{adj}(s) = \phi_{orig}(s) + \frac{s \times 180^\circ}{\lambda_0} \tag{6.26}$$

where the original simulated phase is $\phi_{orig}(s)$, the separation distance is s and the desired, adjusted, simulated phase is $\phi_{adj}(s)$. This correction has been applied to all further Gsolver simulation results.

6.7 RESULTS

This section presents the results of Gsolver rigorous-coupled wave-analysis (RCWA) simulations using the VADR model just described. It is important to note that these simulations are expected to better match the measurements presented in section 5.5.4, because the model has been improved to reflect the lack of anti-reflection coatings on the outside surfaces of the prototype, and to model the loss in silicon substrate.

Since the simulation procedure mimicked the measurement procedure, wherein the TE and TM transmission coefficients were determined separately, the simulation results may be directly compared to the measurement results, as shown in Fig. 6.15. The complex transmission coefficients for the design frequency of 100 GHz are plotted as solid lines for measured data, and as dashed lines for the simulated data, with TE

magnitude in Fig. 6.15(a), TM magnitude in Fig. 6.15(b), TE phase in Fig. 6.15(c) and TM phase in Fig. 6.15(d).

Qualitatively, the agreement between the measured and simulated results is good. The simulations correctly predict the shape of the oscillation in the phase of the TM response, the relative magnitudes of the oscillations in both phase and magnitude responses, and the positions of the peaks and troughs of those oscillations (see Figure 6.15). Quantitatively, the agreement is as follows. The simulations give the TE and TM insertion losses between 6 – 7 dB and 5 — 12 dB respectively, while the TE and TM phases both vary over -22 – -40° and -25 – -75° respectively. Quantitatively, the measured and simulated results for the TM polarisation agree within 1.5 dB and 11° for magnitude and phase respectively, while the TE polarisation shows discrepancies of up to 2.0 dB and 10°. The differences between the measured and simulated data arise from the finite size of the test structure, resulting in field leakage. Also, the slightly spherical nature of the wavefront emitted by the horn antenna may contribute an additional error.

The simulations predict for both TE and TM polarisations a slightly higher insertion loss than was realised in the measurements, which suggests that the fit to the loss of the etalon slightly overstated the loss in the substrate. It would be possible to perform a similar fitting exercise with the VADR simulations to potentially achieve a better fit between the measurements and simulations, however there is little to be gained from this exercise for two reasons. Firstly, the desired outcome has already been achieved. Specifically, the good agreement between these simulations and the measured results confirms that the discrepancies observed earlier was due solely to the original simulations modelling devices with anti-reflection coatings. Secondly, while it is also likely that the lossy substrate has increased the insertion loss of the device, it is not necessary to characterise this absorption loss exactly, because it may be eliminated by fabricating the device from high-resistivity silicon.

It should be noted that a device fabricated with a lossless substrate would still exhibit some loss due to internal reflections, as was demonstrated by the original 1 THz device in the proposal that had 3 dB insertion loss. Due to the limited signal powers exhibited by terahertz sources ($\approx$ 1 mW), in addition to choosing a lossless silicon substrate, it will be necessary to further reduce the insertion loss due to the reflections within the device in order to create a widely useful device. A more suitable, if somewhat arbitrary, figure for the maximum insertion loss is 1 dB. For example, a 1 mW signal impinging on a VADR with 1 dB insertion loss would emerge with 795 μW of signal power, leaving 205 μW to be reflected (assuming a lossless substrate in the device).

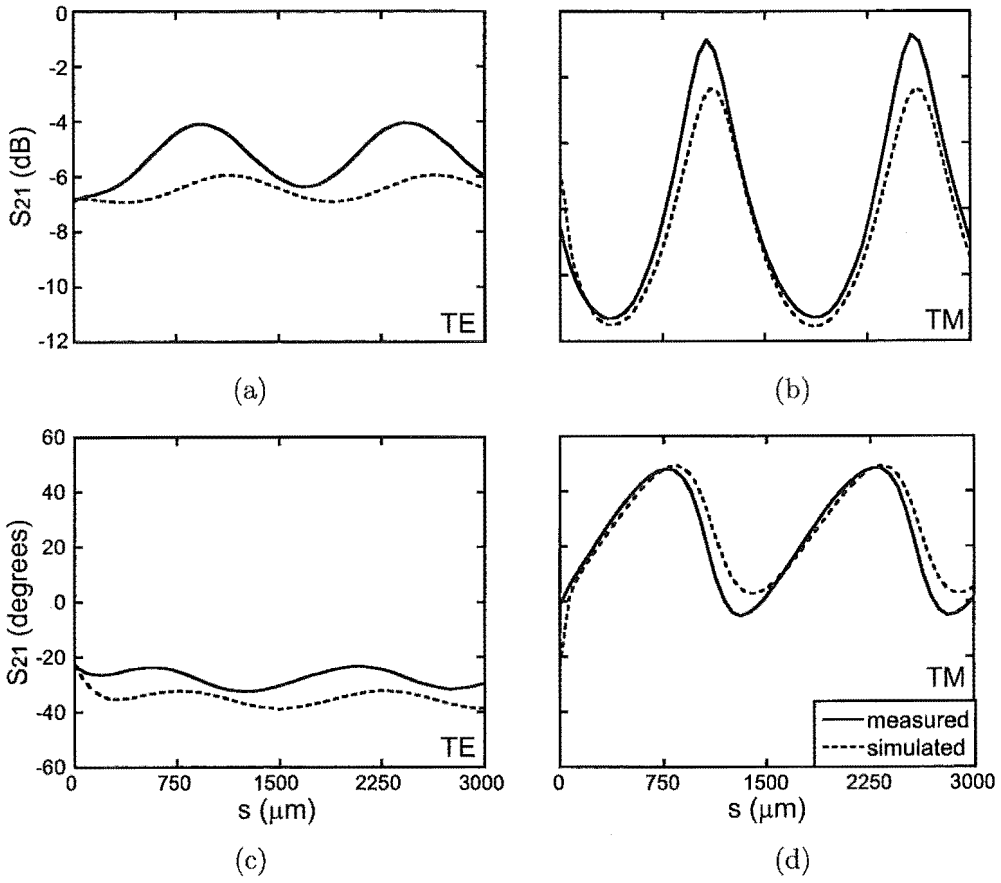


Figure 6.15 The simulated complex transmission coefficients of VADR variable polarisation compensator for incident TE and TM linearly polarised radiation, with measured data for comparison. (a) TE magnitude, (b) TM magnitude, (c) TE phase, and (d) TM phase. Solid lines: measured data with 8.5dB test fixture insertion loss removed. Dashed lines: simulation results

6.7.1 Retardance

The retardance is the key measure of the performance of the VADR device, and the comparison between the measurements and the RCWA simulations is shown in Fig. 6.16. As for the measured results, the retardance is given by the difference in phase between the TE and TM transmission coefficients. According to the simulations, the peak retardance is 82° at a separation of 840μm; with a relative change of 36°, the maximum sensitivity is 360°/mm for separation distances 0 – 100μm (compared to 120°/mm for the measurements), and with a relative change of -33°, the sensitivity is -110°/mm for separations 1000 – 1300μm (compared to -120° for the measurements).

The simulations achieve a greater sensitivity in the interlock region because a zero retardance is achieved. Being able to achieve a zero retardance would be important in practice, both for the increased sensitivity that would result and also so that the device would not have to be removed if it was desired to temporarily leave the incident beam unaltered, for example during an experiment using a variety of polarisations. In the separation region, the simulated sensitivity is slightly lower than that measured, by 10°/mm.

The slight discrepancies between the simulated and measured results are due to the errors already identified in the individual TE and TM transmission coefficients. Overall, there is good agreement and confidence may be had in the Gsolver simulation technique. This is important, because the insertion losses realised to date are too high as already discussed ($> 1\ dB$), and the maximum retardance is too low to be widely useful (less than a quarter wave, $\Gamma < 90°$).

6.8 SUMMARY

It was shown in this chapter that the Gsolver rigorous coupled-wave analysis (RCWA) tool was capable of accurately simulating structures such as VADR with sub-wavelength features. A 10-layer stepped, stair-cased, VADR model was developed that for the first time included the silicon-air interface at the outer surfaces of the device. These were omitted from the original simulations in the proposal because it was assumed that anti-reflection coatings would be used on these surfaces, eliminating reflections. In addition, the complex refractive index of the 10 – 20 Ω-cm silicon substrate (n_s) in the measured VADR device was inferred from a least-mean-squared fit to the measurements of an etalon formed with two plain plates, giving $n_s = 3.41 + 0.257i$. It was shown that this value for the refractive index fell within the upper and lower bounds obtained from the literature for 9.0 Ω-cm and 50 Ω-cm silicon, and that the measured performance of the etalon also fell between the simulated performance of etalons constructed from silicon with these resistivities. A value for 20 Ω-cm silicon was not found.

The results of the RCWA simulations were in good agreement with measurements,

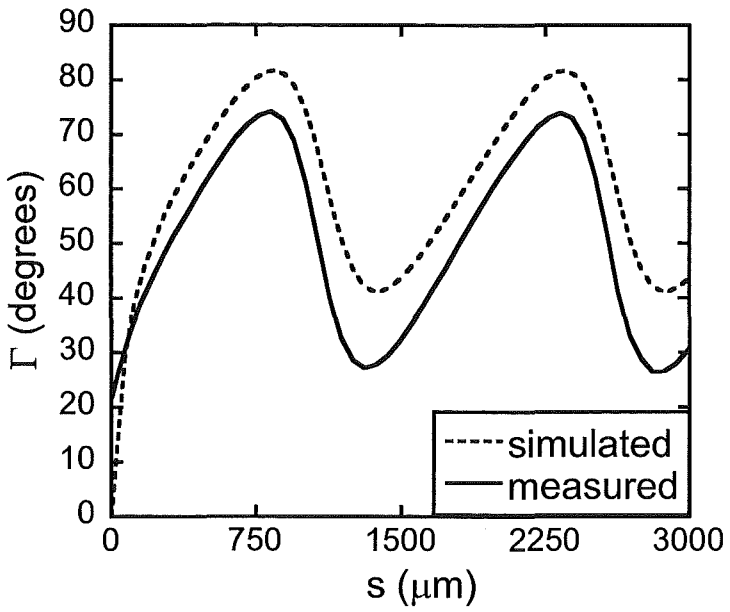


Figure 6.16 Simulated VADR retardance as a function of plate separation, $\Gamma(s)$, with measured data for comparison. Solid line: measured data. Dashed line: simulation results

to within 2 dB and 11° for the magnitude and phase respectively. The good agreement achieved with the improved VADR model confirm that the discrepancies observed between the measured results in Chapter 5, and in the simulations of the original proposal are due to the omission of the anti-reflection coatings in the measured device. They also show that the device still works with a lossy substrate, although it is expected that better performance would result from the use of a lossless substrate such as high resistivity silicon. The good agreement obtained gives confidence in the further use of the Gsolver RCWA tool, and this is important because it is necessary to investigate improved devices that have reduced insertion loss and increased maximum retardance.

Chapter 7

IMPROVED VADR DEVICES

7.1 INTRODUCTION

This chapter describes how the VADR device may be improved. Although there is no particular application in mind now, it is desired to design a device with an insertion loss less than 1 dB and with a maximum retardance in excess of a quarter-wave ($\Gamma = 90°$), or even a half-wave ($\Gamma = 180°$). It is also desired to design a device that is capable of achieving zero retardance ($\Gamma = 0°$) without requiring full interlock between the plates. This is to allow a device to achieve zero retardance in practice, something that is difficult to achieve with the existing device.

To these ends, the addition of quarter-wave anti-reflection coatings are investigated, as are new device geometries. Since it is desired to significantly reduce the insertion loss, it is a requirement of this section that the devices be constructed from lossless or near lossless silicon. Therefore, all simulations in this section assume lossless silicon and use a purely real refractive index.

7.2 ADDITION OF ANTI-REFLECTION COATINGS

In the original proposal, anti-reflection (AR) coatings were assumed to be present on the device [65]. This had the effect of simplifying the simulation model, since the rear surface of the device did not cause reflections and therefore did not need to be included in the model. However, AR coatings were omitted in the measurements and simulations presented in Chapters 5 and 6 respectively, since a suitable coating process had yet to be developed. It was shown in Chapters 5 and 6 that both the maximum retardance and the insertion loss increased when the AR coatings were omitted, although some of the increase in the insertion loss was due to loss in the substrate. In order to judge the role of AR coatings in the apparent trade-off between maximum retardance and minimum insertion loss, it is necessary to simulate a lossless VADR device both with and without AR coatings, which has not yet been done.

In this section, the AR coatings are modelled directly rather than making the

assumption that there simply are no reflections from the rear surface. This is intended to confirm that the original assumption of no reflections from the AR-coated rear surfaces was correct, and that the RCWA tool is in agreement with the MMP tool as to the performance of just such as device. The MMP tool will not be used directly in this work, since the results from the previous investigation have been published [65].

7.2.1 AR coating model

The quarter-wave coatings are designed to eliminate reflections from the unpatterned rear surfaces of the VADR device. At an interface between two regions of different refractive index, such as the example in Figure 7.1, there is generally a non-zero reflection coefficient, T_r , which is given for normal incidence by

$$T_r = \frac{n_s - n_i}{n_s + n_i} \tag{7.1}$$

where n_s, n_i are the refractive indices of the substrate and the incident medium respectively. For a silicon-air interface ($n_s = 3.42, n_i = 1$) such as that found at the rear surfaces of the VADR device, these reflections are significant ($T_r = 0.55$). The fraction of power reflected from just one of these interfaces is equal to $T_r^2 = 0.29$, or 29%. This high level of reflectivity is undesirable because it increases the insertion loss.

In order to eliminate these reflections at a particular frequency, a quarter-wave AR coating may be applied to substrate surface, as shown in Fig. 7.2. The refractive index of the coating, n_q, should be the geometric mean of the refractive indices of the substrate and incident medium:

$$n_q = \sqrt{n_s n_i}. \tag{7.2}$$

For a silicon-air interface, the required refractive index is $n_q = 1.85$. A suitable candidate material for this AR coatings is SU-8 photo resist, with a refractive index of 1.8 at 100GHz [183]. The coating should be one quarter wavelength thick, calculated using the effective wavelength in the coating material,

$$d_q = \frac{\lambda_0}{4n_q}, \tag{7.3}$$

where d_q is the thickness of the coating, and λ_0 is the free space wavelength. At 100 GHz, $\lambda_0 = 3$ mm, and therefore the coating thickness is $d_q = 405$ μm. SU-8 is routinely spin-coatable in layers of up to 1 mm thick [184]. The transmission and reflection intensity of such a quarter wave coating on top of silicon substrate are plotted in Figure 7.3, and they show that the reflected power reduces to zero at 100 GHz. While the reflections are never worse than for an uncoated interface, at other frequencies the reflections are not fully suppressed. Therefore, quarter-wave

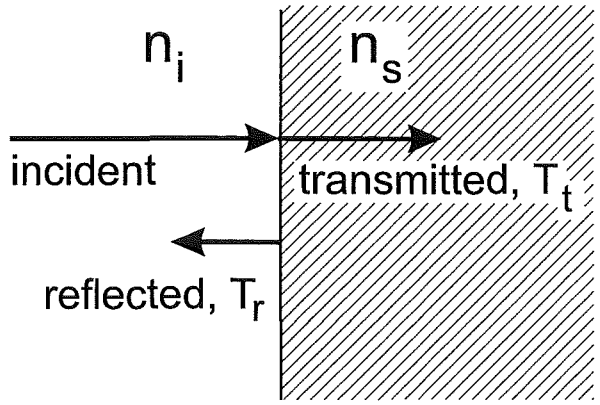


Figure 7.1 Reflections at an interface between two materials of dissimilar refractive index (n_i, n_s). The beam is normally incident on the interface. T_r is the reflection amplitude coefficient and T_t is the transmission amplitude coefficient.

AR coatings are most effective for devices operating at one frequency only. Note that for broad-band operation, an artificial dielectric anti-reflection surface would be ideal, although no such surfaces are explored here.

7.2.2 VADR with AR coatings

Simulations were performed of the VADR with and without the AR coatings. The comparison between non-AR- and AR-coated devices is made here for lossless substrates, to show that the improvement is generally applicable, and not restricted to lossy substrates.

The simulation model for the device with no coatings was the same as in Chapter 6, except that the refractive index of the silicon is now purely real (lossless), $n_s = 3.42$. A VADR device with quarter-wave AR coatings is illustrated in Fig. 7.4, although the stair-casing of the grooves that is employed in the simulation model is not shown. In the simulation model, the coatings are simply entered as uniform layers having refractive index $n_q = 1.84$ and thickness $d_q = 405$ μm.

The simulation results are plotted in Fig. 7.5, with the TE magnitude in Fig. 7.5(a), the TM magnitude in Fig. 7.5(b) and the retardance in Fig. 7.5(c). The solid line is for the devices with AR coatings, while the dashed line is for the device without. It is shown that a VADR device with a lossless silicon substrate, $n = 3.42$, exhibits an insertion loss of 0 – 4 dB for TE and 0 – 10dB for TM. With the addition of the ideal AR-coatings, the insertion loss reduces to 0 – 3 dB for TE and 0 – 4 dB for TM. However the peak retardance falls from 88° to 50°, and the sensitivity falls from 88° / 895 μm (98°/mm) to 50°/845 μm (59°/mm).

The improvement in the insertion loss is most significant for the TM polarisation,

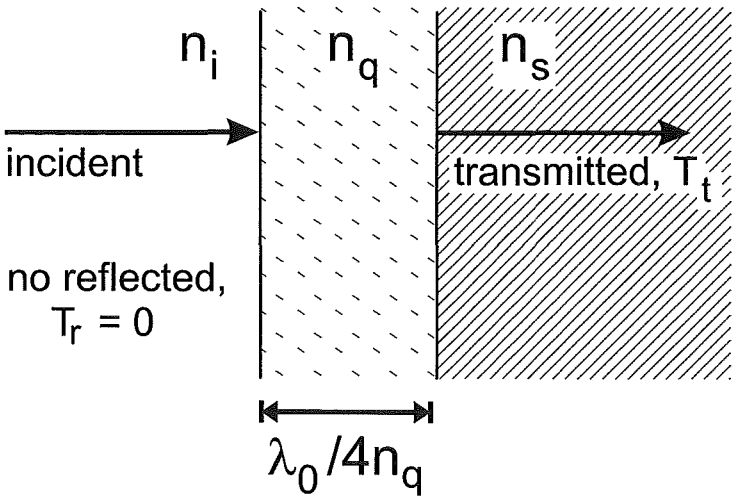


Figure 7.2 A quarter-wave anti-reflection coating may be applied to the substrate to eliminate reflections.

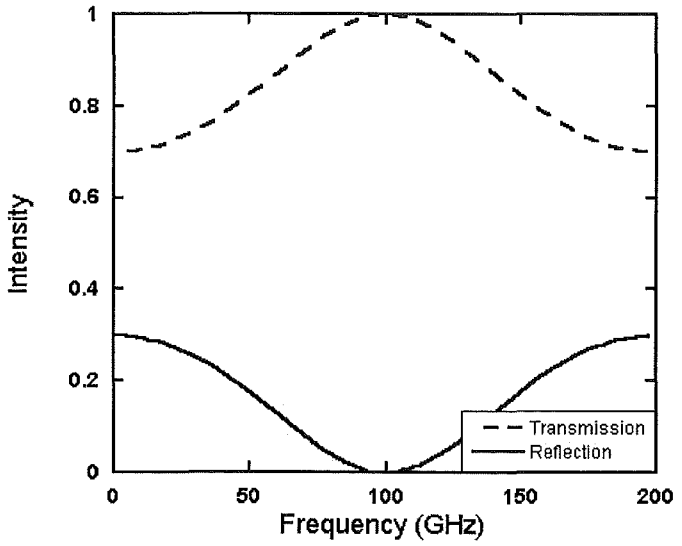


Figure 7.3 The transmission and reflection intensity for a quarter wave coating on a silicon substrate in air. The coating has refractive index n = 1.85, and thickness 405 μm.

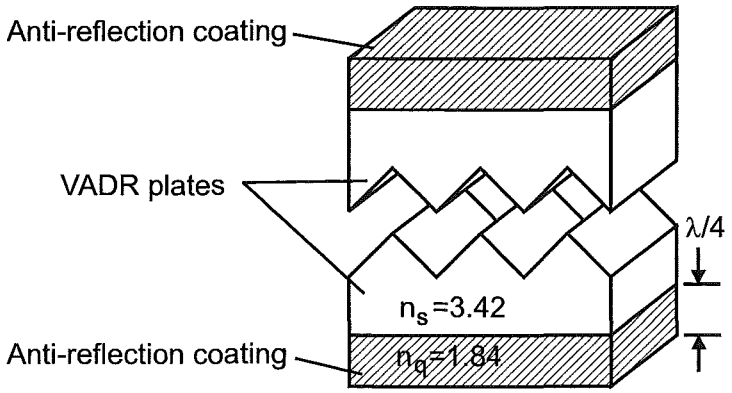


Figure 7.4 VADR with anti-reflection coatings.

reducing the worst loss from 10 dB to 4 dB. The clear improvement in the insertion loss alone recommends the use of AR coatings. Usefully, the difference between the TE and TM magnitudes is reduced as well, with both being in the range 0 – -4 dB. This is important because it results in less rotational error being introduced to the transmitted polarisation during operation.

Unfortunately, the trade-off in using the AR coatings is that the maximum retardance is reduced. Since VADR without AR coatings does not quite reach a quarter-wave retardance, improvements would have been required in any case. The much reduced retardance in the device with coatings indicates that it will be necessary to investigate alternative grating geometries in order to find a structure that has enough birefringence to develop a quarter-wave or half-wave retardance once the AR coatings have been applied. Thus, all new devices in this chapter include AR-coatings.

7.3 ALTERNATIVE GEOMETRIES FOR IMPROVED PERFORMANCE

The desirable performance characteristics of a variable polarisation compensator are low insertion loss, a highly sensitive, linear variation in retardance as a function of plate separation R(s), and a peak retardance in excess of 90° (quarter-wave device), or 180° (half-wave device). A 90° retardance could be produced by cascading two VADR devices in series, and 180° by three. For practical reasons, it is preferable to use only a single device or reduce the number of devices in the cascade to two, by using devices with a greater retardance.

For this reason, alternative grating geometries were investigated that offered greater birefringence. Simulations are presented of both single and double grating-pair devices with rectangular grooves. Three devices are proposed, and examples of each are simulated. The three devices are shown in Fig. 7.6, and are named as follows: the R-VADR

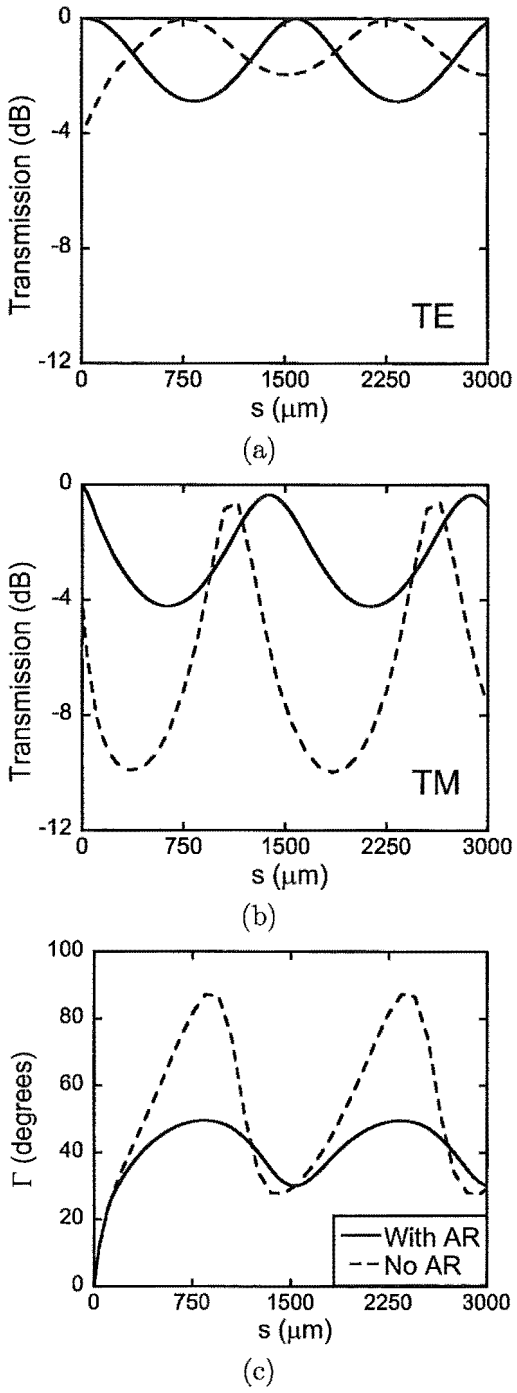


Figure 7.5 Simulated VADR performance with and without anti reflection coatings at 100 GHz. (a) TE magnitude, (b) TM magnitude, (c) retardance, $\Gamma(s)$.

is a single rectangular grating device, shown in Fig. 7.6(a); the R^2-VADR has two sets of gratings, shown in Fig. 7.6(b); the The dual axis rectangular tooth half-wave VADR (DARTH-VADR) is an R^2 with orthogonal gratings, as shown in Fig. 7.6(c).

It is predicted that the R-VADR is capable of a quarter wave retardance, the R^2-VADR a half wave retardance, while the DARTH-VADR achieves a both positive and negative quarter-wave retardances (for a total range of a half-wave), and has the advantage of not requiring full interlock for any retardance in that range. These rectangular groove devices are found to produce enhanced sensitivity and reduced loss compared to the equivalent VADR structures. Hence, no further simulations are presented for V-grooved devices.

7.4 R-VADR

The R-VADR differs from the VADR only in that the grooves are rectangular, as shown in Fig. 7.6(a). The basic theory of operation remains the same, but the retardance is increased because the rectangular grooves provide a greater birefringence.

The rectangular cross-sectional geometries required for R-VADR can be made using deep reactive ion etching (DRIE) micromachining techniques for short operating wavelengths where groove depths up to approximately 300μm are required. For deeper grooves at short operating wavelengths, excimer laser ablation may be used; for longer operating wavelengths, the lateral feature sizes are large enough for mechanical cutting techniques to be used. All three techniques allow the depth to be controlled precisely, and the groove depth to be chosen to be any value without the restrictions imposed by the KOH etch technique. Thus, devices of the same 500μm period, but with three different groove depths (g), (see Fig. 7.6a), have been investigated: the direct equivalent to the VADR ($g = 350\mu$m), one shallower ($g = 200\mu$m) and one deeper ($g = 500\mu$m). The same RCWA technique was used for these simulations as before. The retardance for all three R-VADR devices is plotted in Fig. 7.7(a), while the magnitude response for only the 350μm R-VADR is plotted in Fig. 7.7(b).

Examining Fig. 7.7(a), it is clear that g plays a pivotal role in determining the peak retardance. We find that the greater g, the greater the peak retardance. The specific retardances and total sensitivities are recorded in Table 7.1 for each device. The

The 350μm R-VADR gains 49° in peak retardance and 60°/mm in sensitivity, compared to the simulations for the improved 350μm VADR.

The magnitude response of the 350μm R-VADR (see Fig. 7.7b) indicates a reduction in the insertion loss of 1dB, compared with the equivalent VADR (Fig. 7.5), at the expense of asymmetry. The insertion loss oscillates from 0 – 0.3dB(TE) and 0 – 0.6dB (TM) in the interlock region, and 0 – 2.6dB (TE) and 0 – 2.0dB (TM) in the separation region, showing an asymmetry of 0.6dB between the maximum losses in TE

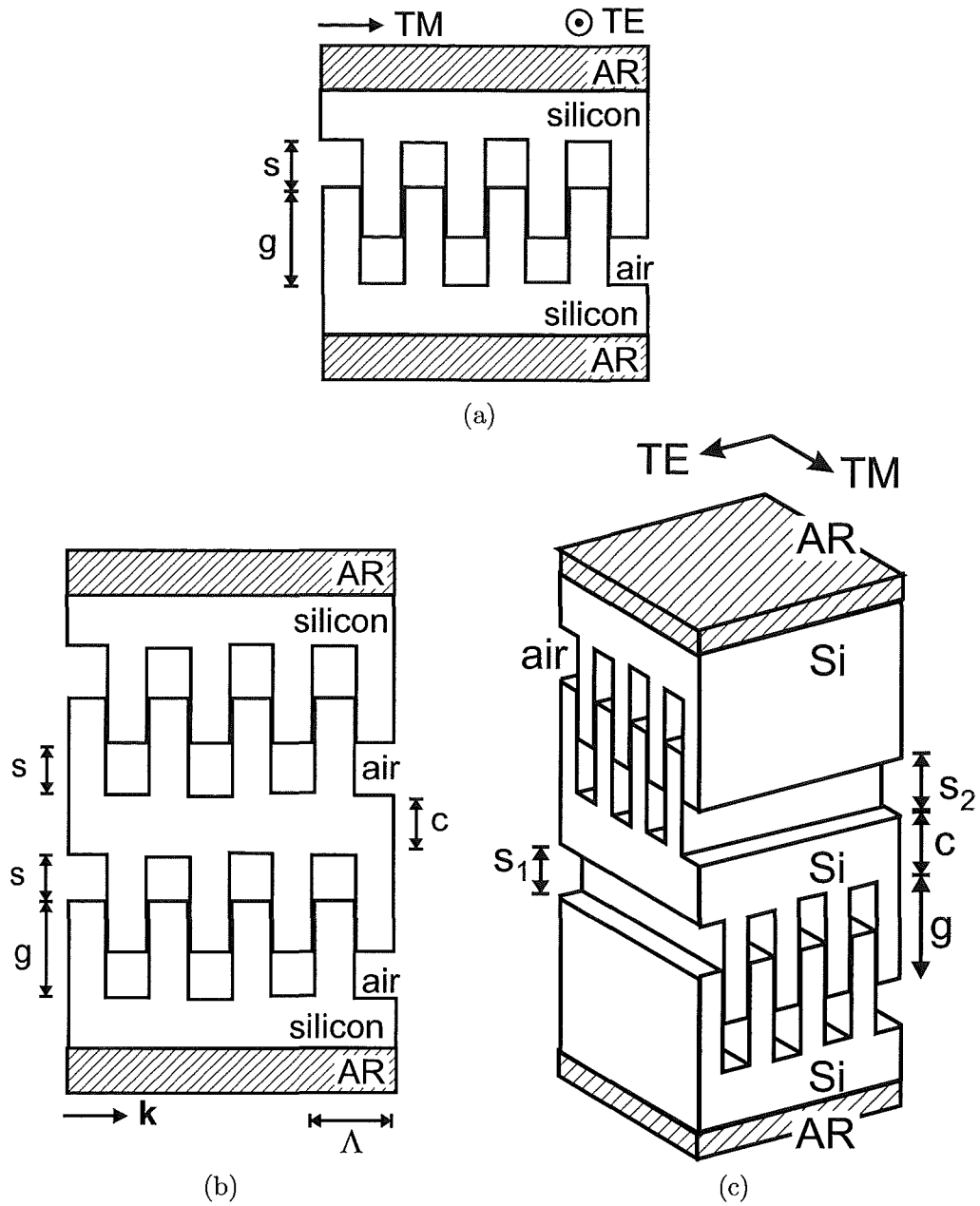


Figure 7.6 The three new VADR devices. The dimensions are groove depth g, separation s and centre-thickness c. (a) R-VADR, (b) R^2-VADR, (c) DARTH-VADR.

Table 7.1 Characteristics of the R-VADR retardance $\Gamma(s)$ for various groove depths at 100 GHz.

Groove depth g (μm)	Peak retardance (degrees)	Separation at peak retardance (μm)	Total sensitivity (degrees/mm)	Insertion loss (TE, dB)	Insertion loss (TM, dB)
200	54	800	67	0 – 2.8	0 – 3.6
350	112	750	150	0 – 2.6	0 – 2.0
500	147	850	173	0 – 5.0	0 – 5.0

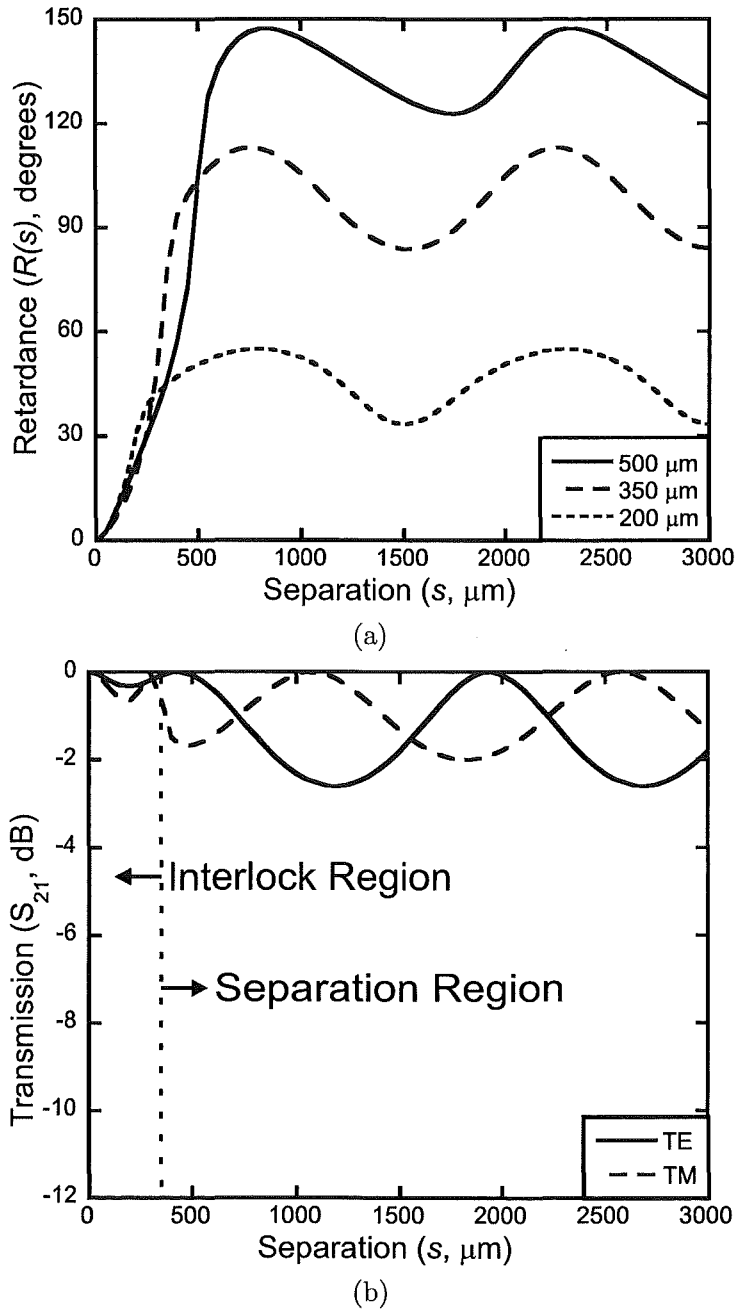


Figure 7.7 Transmission characteristics of R-VADR variable polarisation compensators, (a)retardance $R(s)$ for three devices with different groove depths and (b) TE and TM insertion losses for a 350μm groove depth R-VADR.

and TM.

As with the VADR, a distinct transition is seen from the interlock region into the separation region. The interlock region retardance is characterised by a steep increase; the separation region by a steady oscillation. The transition is also clear in the magnitude responses of Fig. 7.7(b).

The R-VADR is easily capable of achieving a quarter-wave retardance, with an acceptable magnitude response. The improved performance is attributed to the higher total birefringence of the rectangular gratings that now have a constant effective dielectric constant as a function of depth, rather than being smoothly graded as in the V-groove case. The 500μm groove depth on a 500μm period corresponds to a 2:1 aspect ratio, which is at the limit of DRIE fabrication technology.

7.5 R^2-VADR

A device capable of a half-wave retardance makes an attractive proposition, as this is necessary for complete polarisation control. Aspect ratio limitations in the fabrication process (that limit groove depths) may be overcome by coupling two R-VADRs in series, combining the inner two plates into a double-sided grating, as shown in Fig. 7.6(b). This device is named the double rectangular groove VADR, or R^2-VADR. The behaviour of an R^2-VADR is determined by both the groove depth, g, as for the R-VADR, but also by the thickness of the substrate between the centre plate's two gratings; this is the dimension c in Fig. 7.6(b).

The substrate thicknesses on the outer plates have no effect on the performance, as the outer surfaces are AR coated. To adjust the separation distance of the R^2-VADR, the centre plate is held fixed, while the two outer plates are moved by the same distance s, such that they are always equidistant from the centre plate. Alternatively, asymmetrical operation of this device is possible, as there are now two degrees of freedom for plate separation, but these cases are not explored here.

7.5.1 T-matrix simulation results

The full operation of 10,000 separate devices, in positional settings 0 - 3mm, was simulated using the transfer matrix code and Rytov's exact EMT (see section 2.4.4), covering all combinations of c and g from 10μm to 1mm in 10μm steps. To aid the selection of top performing devices, the transmission coefficient data from each simulation was post processed to provide two performance characteristics: peak retardance R_{max}, and average insertion loss L_{av}. These are defined as

$$R_{max} = \arg\Big(t_{TM}(a)\Big) - \arg\Big(t_{TE}(a)\Big) \tag{7.4}$$

$$L_{av} = \frac{1}{b}\int_0^b \Big(|t_{TM}(x)| + |t_{TE}(x)|\Big)dx \tag{7.5}$$

where x is the separation distance, a is the separation distance at which the retardance is a maximum, b is the separation distance at which the retardance first reaches 180°, t_{TM} and t_{TE} are the complex transmission coefficients for TM and TE polarisations respectively, and arg(t) is the phase of the complex transmission coefficient t.

The performance characteristics were calculated from the results of each device in the simulation, and they are shown in the contour plots of Fig. 7.8. The contour plots clearly show that the influence of the c and g dimensions on device performance. Figure 7.8(a) shows that the peak retardance R_{max} is determined primarily by the groove depth, g, although the choice of c is also important. As for the R-VADR, we find that the greater g, the greater the peak retardance, with full 360° retardance possible for devices with $g \geq 600$ μm. Figure 7.8(b) shows a pattern of peaks and troughs in the average insertion loss L_{av}, with periodic dependence on both c and g. The periodicity corresponds to a half wavelength of the incident radiation in silicon. Further simulations of devices with c and g up to 3mm, not shown here, confirm the continuation of this periodic pattern. It generally desired to have a device with high peak retardance R_{max}, and low average insertion loss L_{av}. In the both plots, the better the performance the lighter the colour. Given a specific retardance requirement, an approximate minimum groove depth, g_{min}, can be obtained from Fig. 7.8(a). Then, values must be found for c and $g > g_{min}$ that fall within a white region on Fig. 7.8(b).

7.5.2 Characteristics of an example R^2-VADR device

An example device was selected from the domain shown in Fig. 7.8 to have the greatest ratio of R_{max}/L_{av} . The device is marked with a + in both parts of Fig. 7.8, had dimensions $c = 840\mu$m and $g = 730\mu$m, and performance characteristics $R_{max} = 440°$, $L_{av} = 0.15$dB. Other performance criteria could be used to select devices depending on the relative importance of the retardance and loss. The characteristics of this 50μm period R^2-VADR were determined using the transfer matrix code, and are plotted as solid lines in Fig. 7.9, with retardance $R(s)$ in Fig. 7.9(a), TE magnitude in Fig. 7.9(b) and TM magnitude in Fig. 7.9(c). The peak retardance is 446°, at a separation of 1030μm. We define the operating range for the device as retardances from 0 – 180°, which correspond to separations of 0 - 290μm. Over the operating range, the sensitivity is 620°/mm, four times greater than the 350μm groove depth R-VADR (Fig. 7.7). The insertion losses are low, from 0 – 0.2dB at separations 0 – 290μm for TE, and 0 – 0.1dB at separations 0 – 260μm, increasing to 1.3dB at 290μm for TM. The maximum insertion loss in the operating region is 1.3dB better than the 350μm R-VADR, although the asymmetry is 0.5dB larger, at 1.1dB. Outside of the operating region, the retardance

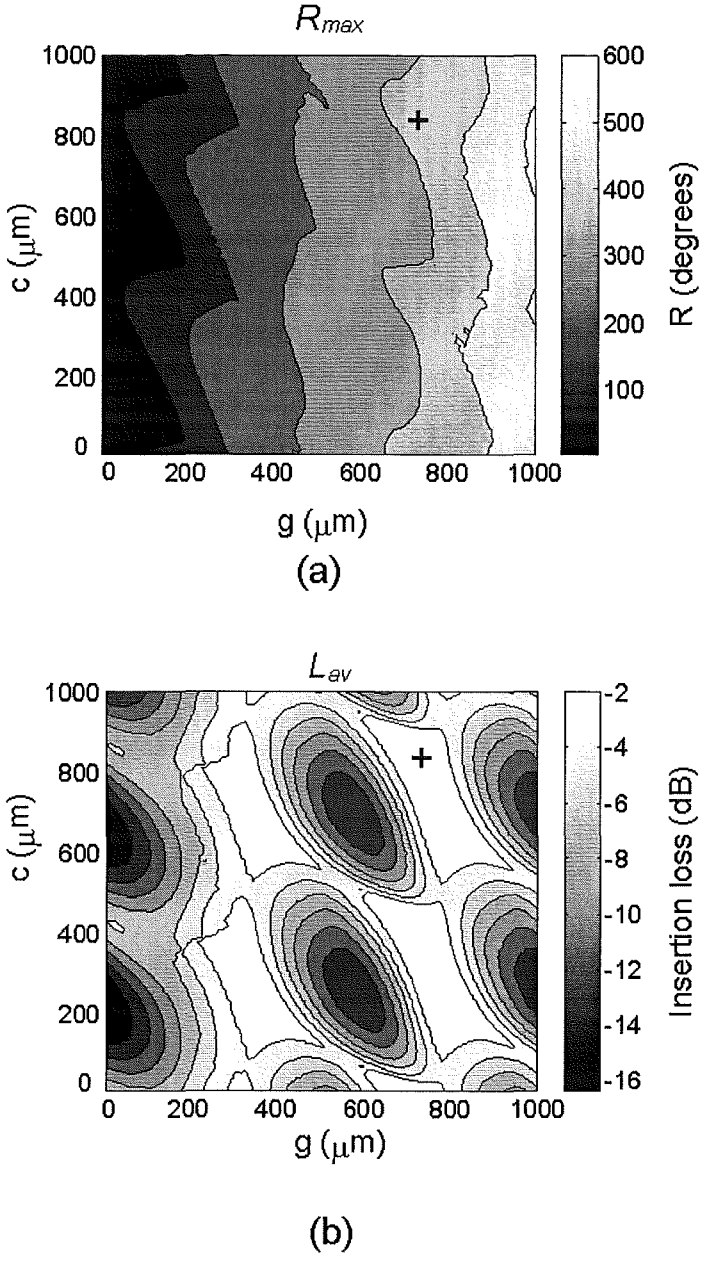


Figure 7.8 Contour plot analysis of 10,000 different R^2-VADR devices with varied centre thickness c and groove depth g; (a) peak retardance R_{max}, (b) average insertion loss L_{av}.

oscillates from 335 - 445°, for insertion losses of 0 – 10dB for TE and 0 – 17dB for TM. These large insertion losses outside the operating region are not unexpected, as it was only the loss in the region of small plate separations that was used as one of the device selection criteria.

RCWA simulation results for the same R^2-VADR are plotted as solid circles in Fig. 7.9, and show excellent agreement. This device would be impractical to fabricate (50μm period and 730μm groove depth) so RCWA simulation results for an R^2-VADR with the same c and g dimensions, but with a 500μm period are plotted as open squares in Fig. 7.9. The peak retardance is reduced by 40° to 416°, although it occurs at the same separation of 1030μm. The sensitivity over retardances 0 – 180°, which correspond to separations 0 – 370μm, is reduced by 135°/mm to 485°/mm. The operating region insertion losses are improved, at 0 – 0.1dB for TE and 0 – 0.1dB, except for a 0.6dB peak at 300μm separation, for TM. The asymmetry is further reduced to 0.6dB. Outside the operating region, the insertion loss is 0 – 10dB for TE and 0 – 5dB for TM. The difference in performance between the 50μm and 500μm period devices is attributed to the onset of diffraction in 500μm period device, since $\lambda_{Si}/2 < d$, where $\lambda_{Si} = \lambda_0/n_{Si} = 877\mu m$ and d is the period. Despite the differences, the simple transfer matrix simulations provide a useful tool for confidently estimating zones of comfort for good performance with large peak retardance and low insertion loss. It is accurate for small period structures, but the RCWA method is required to get detailed performance characteristics for larger period devices.

The excellent insertion loss figures for the R^2-VADR will be subject to increase in a practical device since there will be additional loss owing to absorption in the substrate, but this can be minimised by using high resistivity wafers [16]. Other practical considerations include the possibility that mechanical imperfections may prevent full interlock. Additionally, any clearance between the grooves to alleviate stiction would result in a slight, but acceptable, performance degradation. It is unfortunate that in both cases, the ability of the device to reach zero retardance would be compromised.

7.6 DARTH-VADR

It is difficult to control a process such that two separate surfaces, fabricated at separate times, can be perfectly interlocked. A device configuration that could tolerate such imperfections, yet still give 0° retardance is necessary for this to be a truly practical device. The benefits of low loss, high sensitivity and a maximum retardance in excess of 90° as obtained with the R^2-VADR must be retained in order to prove that all areas of operation can be optimised together.

The Dual Axis Rectangular Tooth Half-wave VADR (DARTH-VADR) achieves these goals. The design of the device is illustrated in Figure 7.6(c). The DARTH-VADR is similar to the R^2-VADR, except that the two pairs of grooves are rotated so

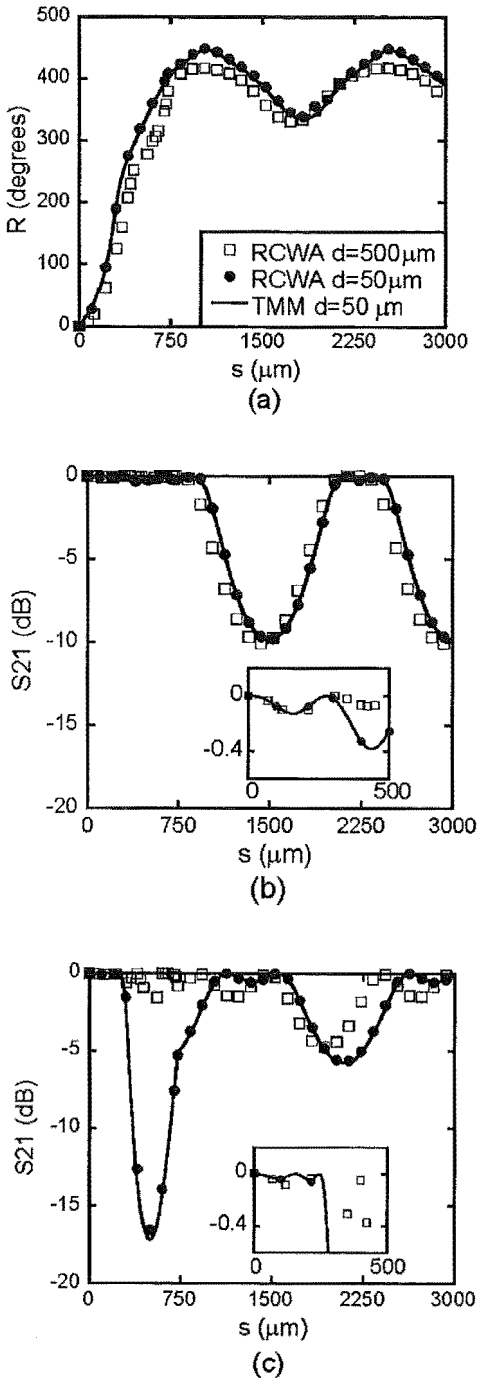


Figure 7.9 Transmission characteristics of R^2-VADR devices with dimensions c=840μm, g=730μm: (a) retardance $R(s)$, (b) TE insertion loss, (c) TM insertion loss. Solid lines and symbols: simulations for 50μm period devices. Open symbols: 500μm period device. Insets in (b) and (c) show insertion loss behaviour for small separations.

that they are orthogonal. In the initial position, with the central plate centred, the polarisation rotation from one device is offset by the polarisation rotation from the other device, giving an overall zero change in the polarisation state. To change the polarisation, the central plate is moved to one side or the other. This device is capable of giving -90 – 90° of retardation (this is a range of a half wave, and hence the "half wave" in the device's name).

7.6.1 T-matrix simulation results

As for the R^2-VADR, T-matrix simulations were used to estimate zones of comfort for good performance. For the DARTH-VADR the outer two plates are fixed while the centre plate is moved. Here the separation is defined with respect to the upper set of gratings, $s = s_1$. The two outer plates were positioned so that $s_1 + s_2 = g$, according to the requirement the grooves remain interlocked in all positions. Therefore, the range of separations for any particular device was $0 \leq s \leq g$ where g is the groove depth defined in Fig. 7.6(c).

The full operation of 10,000 separate DARTH-VADR devices, in positional settings $0 - g$, was simulated using the transfer matrix code, covering all combinations of c and g from 10μm to 1mm in 10μm steps.

To aid the selection of top performing devices, the transmission coefficient data from each simulation was post processed to provide two performance characteristics: peak difference between positive and negative retardance ΔR_{max}, and average insertion loss L_{av}. These are defined as

$$R_{max} = (\arg(t_{TM}(a_1)) - \arg(t_{TE}(a_1))) - (\arg(t_{TM}(a_2)) - \arg(t_{TE}(a_2))) \tag{7.6}$$

$$L_{av} = \frac{1}{b_1 - b_2} \int_{b_2}^{b_1} (|t_{TM}(x)| + |t_{TE}(x)|)\, dx \tag{7.7}$$

where x is the separation distance, a is the separation distance at which the retardance is a maximum (a_1) or a minimum (a_2), b is the separation distance at which the retardance first reaches +90° (b_1) or -90° (b_2), t_{TM} and t_{TE} are the complex transmission coefficients for TM and TE polarisations respectively, and arg(t) is the phase of the complex transmission coefficient t.

The performance characteristics were calculated from the results of each device in the simulation, and they are shown in the contour plots of Fig. 7.10. The c and g dimensions may be seen to strongly influence the device performance. It may be seen in Fig. 7.8(a) shows that the peak retardance R_{max} is determined primarily by the groove depth, g. Unlike the R^2-VADR, the choice of c appears to be unimportant. As

before, it is found that the greater g, the greater the peak retardance, both positive and negative. A half-wave range of retardances ($|\Delta|Gamma_max| \geq 180°$) is possible for devices with $g \geq 310$ μm. Fig. 7.8(b) shows a pattern of peaks and troughs in the average insertion loss L_{av}, with periodic dependence on both c and g. However, there the groove depth g appears to cause a greater variation in the average loss than the centre thickness c. As before, the periodicity corresponds to a half of the material wavelength of the 100 GHz illumination ($\lambda_n/2 = 439$ μm).

Since it is generally desired to have a device with high peak retardance R_{max}, and low average insertion loss L_{av}, both these values may be taken into account when choosing an example DARTH-VADR. In the both plots, the better the performance the lighter the colour. Given a specific retardance requirement, an approximate minimum groove depth, g_{min}, can be obtained from Fig. 7.10(a). Then, values must be found for c and $g > g_{min}$ that fall within a white region on Fig. 7.10(b). In order to make the selection of the centre thickness c more accurate, an expanded view of one of the regions of low loss is presented in Fig. 7.10(c).

7.6.2 Example device

An example device was chosen from the domain shown in Fig. 7.10 to have the greatest ratio of $|\Delta\Gamma_{max}|/L_{av}$. The device is marked with a + in all three parts of Fig. 7.10, and had dimensions $c = 200$ μm and $g = 390$ μm. The performance characteristics where $|\Delta\Gamma_{max}| = 240°$, $L_{av} = 0.25$ dB. Other performance criteria could be used to select devices depending on the relative importance of the retardance and loss. The characteristics of this 50μm period DARTH-VADR were determined using the transfer matrix code, and are plotted as solid lines in Fig. 7.11, with retardance $R(s)$ in Fig. 7.11(a), TE magnitude in Fig. 7.11(b) and TM magnitude in Fig. 7.11(c). The peak retardances are +120 ° and -120°, at separations of 0 μm and 390 μm respectively. The operating range of the device is defined as retardances from -90 – 90°, which correspond to separations of 36 – 352 μm. Over the operating range, the sensitivity is 180°/316 μm (570 °/mm), almost four times greater than the 150°/μm sensitivity of the 350μm groove depth R-VADR (Fig. 7.7). The insertion losses are low, at less than 0.6 dB throughout the operating range (separations 36 – 352 μm). The separation at zero retardance is 195 μm, which is exactly half the groove depth. The maximum insertion loss in the operating region is 2 dB better than the 350μm R-VADR, although the asymmetry is similar, at 0.6dB. Outside of the operating region the insertion loss increases to 2 dB. These slightly larger insertion losses outside the operating region are not unexpected, as it was only the loss in the region of small plate separations that was used as one of the device selection criteria.

RCWA simulation results for the same DARTH-VADR are plotted as solid circles in Fig. 7.11, and show excellent agreement. This device would be difficult to fabricate

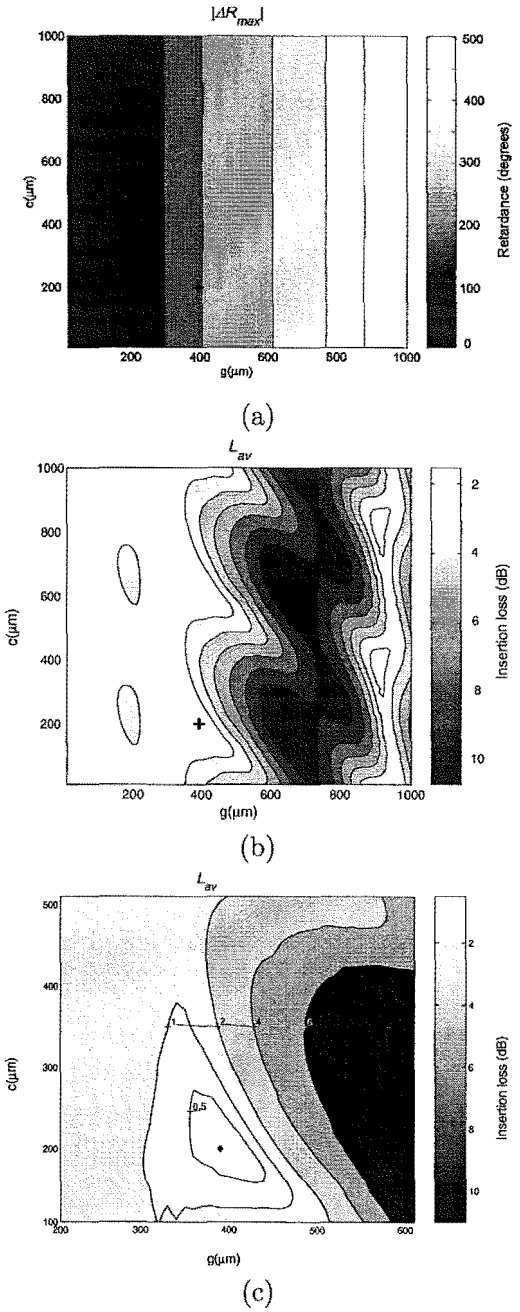


Figure 7.10 Contour plot analysis of 10,000 different DARTH-VADR devices with varied centre thickness c and groove depth g; (a) peak retardance $|\Delta R_{max}|$, (b) average insertion loss L_{av}, (c)L_{av} for a subset of the devices.

(50μm period and 390 μm groove depth) so RCWA simulation results for a DARTH-VADR with the same c and g dimensions, but with a 500 μm period are plotted as open squares in Fig. 7.9. The peak retardance is reduced by 38° to 82°. The sensitivity over retardances -82 – 82°, which correspond to separations 0 – 390 μm, is reduced by 280°/mm to 164°/390 μm, (290°/mm). The operating region insertion losses are improved, with a maximum of 0.4 dB for TE and TM. The asymmetry is further reduced to 0.4 dB. The difference in performance between the 50μm and 500μm period devices is attributed to the onset of diffraction in 500μm period device, since $\lambda_{Si}/2 < d$, where $\lambda_{Si} = \lambda_0/n_{Si} = 877\mu m$ and d is the period. The discrepancies between the exact Rytov EMT and RCWA simulations of the 50 μm period device are attributed to small grating depth errors that are exacerbated by having the gratings crossed.

Despite the differences, the simple transfer matrix simulations provide a useful tool for confidently estimating zones of comfort for good performance with large peak retardance and low insertion loss. It is accurate for small period structures, but the RCWA method is required to get detailed performance characteristics for larger period devices.

The DARTH-VADR overcomes the difficulty of achieving zero retardance in practice, because it does not require full interlock between the plates. Furthermore, the mode of operation investigated here required only one plate to be moved, in comparison to the two plates that needed to be moved for R^2-VADR. Also, the dimensions of the DARTH-VADR plates are much more amenable to fabrication than those of the R^2-VADR, since the groove depth and aspect ratio are readily attained by existing microfabrication processes such as deep reactive ion etching.

7.7 COMPARISON WITH BABINET COMPENSATOR

In order to emphasise the excellent prospects for the DARTH-VADR design, a quarter-wave Babinet compensator is designed for use at 100 GHz. The dimensions of the Babinet compensator are shown in Fig. 7.12. The refractive indices are n_o for the ordinary ray and n_e for the extraordinary ray. No tabulated value is available for the refractive indices of calcite at 100 GHz, but they are available for the nearby 90 GHz (see Table 2.2). This is the only tabulated value in the terahertz region for which calcite is lossless, so it makes for an overly optimistic impression of its performance. Nonetheless, the adopted refractive indices are $n_o = 2.946$ and $n_e = 2.883$.

The maximum wedge thickness may be calculated such that the maximum retardance is a quarter-wave as desired. From section 2.3, the retardance Γ in any setting is given by

$$\Gamma = \Gamma_1 + \Gamma_2 = \frac{2\pi}{\lambda_0}(d_1 - d_2)(|n_o - n_e|), \tag{7.8}$$

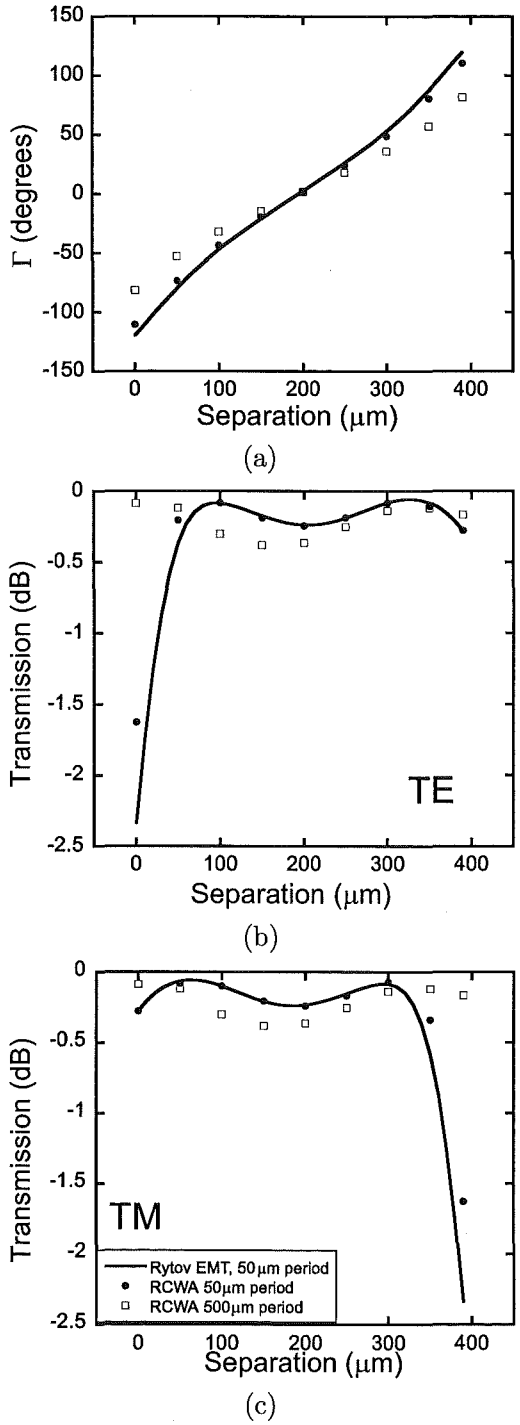


Figure 7.11 Transmission characteristics of a DARTH-VADR device with dimensions c=200μm, g=390μm: (a) retardance $R(s)$, (b) TE insertion loss, (c) TM insertion loss. Solid lines and symbols: simulations for 50μm period devices. Open symbols: 500μm period device. Insets in (b) and (c) show insertion loss behaviour for small separations

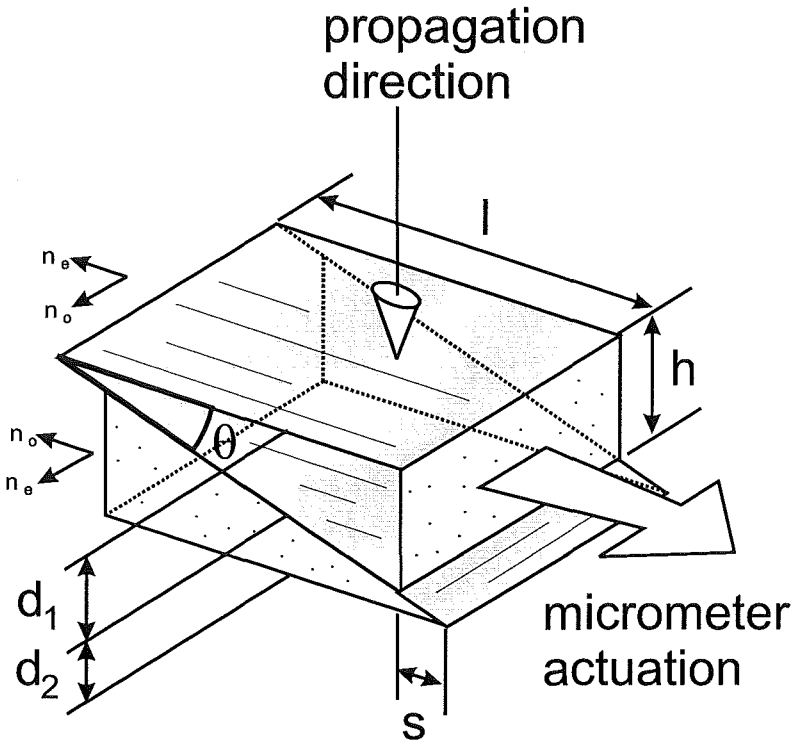


Figure 7.12 Dimensions of the Babinet polarisation compensator. The wedge angle is θ, while the length is l and the height is h. The wedge thicknesses at the beam are d_1 and d_2, while the offset in the relative position of the two wedges is s.

where Γ_1 and Γ_2 are the retardance experienced in the top and bottom wedges, In order to achieve a retardance of 90° ($\pi/2$), the difference in wedge thicknesses at the beam must be

$$d_1 - d_2 = \frac{\Gamma\lambda_0}{2\pi(n_o - n_e)} = 11.9 \text{ mm}. \tag{7.9}$$

The wedges are usually thin, and the typical wedge angle of 2.5° is adopted here. Thus the minimum wedge length may be found by trigonometry to be 272 mm. Since the beam cannot be at the edge of the wedge, the actual wedge length would need to be even larger. In order to move between zero and quarter-wave retardance, the wedge would need to be shifted by 272 mm, giving a sensitivity of 90°/272 mm (0.33°/mm). This less than one hundredth of the sensitivity of the original VADR, and certainly could not be actuated by a compact piezo-electric stack in the same way that VADR could be. This would severely limit its ability to respond to fluctuations in the polarisation state (such as in an application requiring dynamic polarisation compensation).

It is clear that the dimensions required for a terahertz Babinet compensator are prohibitively large. Even if two sufficiently large pieces of calcite could be found, the device would be bulky, clumsy, and adjust slowly. Due to calcite's strong dispersion, the size is not guaranteed to be smaller at higher frequencies, and even so, the device would still be useless because at any other frequency in the terahertz region, the losses in the calcite are high, in some cases rendering it opaque. Therefore, existing optical compensators are not a viable solution for use in the terahertz region, and the VADR devices proposed here offer an excellent alternative.

7.8 SUMMARY

This chapter investigated several improved VADR devices, and found them to be far superior to existing optical techniques, such as the Babinet compensator, at terahertz frequencies. Anti-reflection coatings eliminate reflections at the outer surfaces of the device, which has the advantage of reducing the insertion loss overall, and also reducing the difference between the TE and TM insertion losses. The trade-off is that the maximum retardance is reduced significantly. In order to overcome this reduction, devices with higher birefringence grooves were investigated. The three devices had rectangular grooves, and were named R-VADR for the single grating pair device, R^2-VADR for the dual grating pair device, and DARTH-VADR for the orthogonal dual grating pair device. The R-VADR was capable of a quarter-wave retardance with 2 dB of insertion loss. The R^2-VADR was capable of a half wave retardance with 0.6 dB of loss, while the 50 μm period DARTH-VADR was capable of both a positive and negative quarter-wave shift with 0.6 dB of loss. The DARTH-VADR also overcame the critical difficulty of achieving zero retardance in a practical device. The other designs

(VADR, R-VADR, R^2-VADR) all require the plates to interlock fully and perfectly to give zero retardance, but the difficulty of achieving this was highlighted by the prototype investigated in chapter 5. The DARTH-VADR does not require the plates to interlock, because the orthogonal grating pairs counteract one another giving zero retardance at a separation approximately equal to the half the groove depths. The devices investigated in this chapter may be scaled to operate anywhere in the terahertz region with the same high levels of performance.

Chapter 8

VARIABLE PHOTONIC BAND GAP FILTER

8.1 INTRODUCTION

The variable photonic band gap filter (VPBGF) devices presented in this chapter are inspired by DARTH-VADR's crossed-gratings, interlocking teeth and mechanical actuation schemes. The resulting device is a tunable simple-tetragonal (st) "layer-by-layer" or "woodpile" photonic crystal. The concept of the VPBGF may be illustrated by the two example positional settings of the same device in Fig. 8.1, with the device in a "fully interlocked" position in Fig. 8.1(a) and in a "partially interlocked position" in Fig. 8.1(b). This is a completely new method for making a photonic crystal, that offers a wide range of potential tuning methods. For ultimate control of a woodpile-like crystal it would be necessary to individually actuate each of the rods, but this is not practical. Whereas, the VPBGF (an st-like woodpile) may be reconfigured and tuned in a large number of ways using easily accessible external surfaces (plates). None of the existing stacked plate structures [114, 116, 185, 186] offer the extra degree of freedom provided by the interlocking plates of the VPBGF.

The structure in Fig. 8.1(a) has the appearance of the st woodpile, which is known to have useful properties such as a full 3-D band gap and strong stop band attenuation [186–189], while the structure in Fig. 8.1(b), and a diversity of other structures with different spacings, exhibit a range of defect modes that have not previously been studied, and variations on the one related mode that has been studied (planar air cavity). This diversity of defect modes may be exploited to create a variety of tunable filters.

This chapter studies the effect of the device dimensions on the location of the photonic band gaps in VPBGF devices, and how various actuation methods tune the band gap and associated defect modes. One example is experimentally demonstrated.

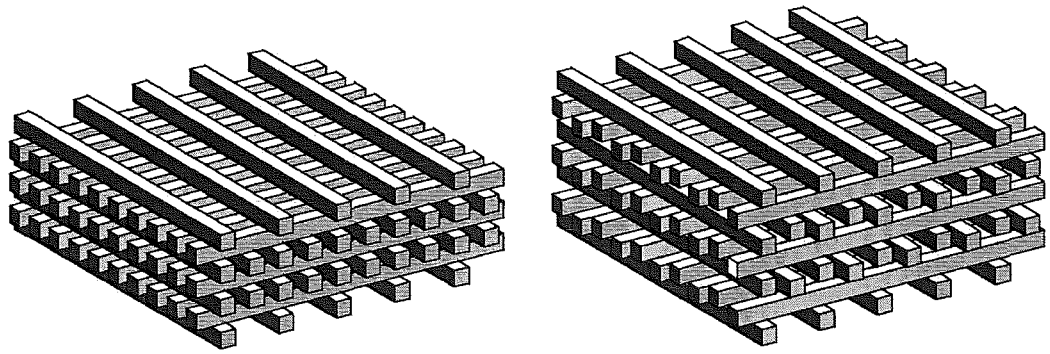


Figure 8.1 The VPBGF actuation method applies to an st woodpile (a) original st woodpile (b) example of one novel expanded position (others are introduced later).

8.2 VPBGF PLATES

The VPBGF device is constructed from a number of plates that each comprise two orthogonal grids, such as the one shown in Fig. 8.2. The plate does not fall apart because the grids are joined together. This may be achieved by fabricating the plates from a solid blank. For example, a silicon plate could be fabricated from a single wafer using micromachining techniques such as two-sided etching.

The dimensions of the plate, and definitions of the separation and shift distances are shown in Fig. 8.3. The period of the rods is denoted by Λ and defined as the in-plane (not stacking) repeat distance of the rods. It should be noted that the period of VPBGF rods in a fully interlocked position is twice as large as the period that would ordinarily be defined for an st woodpile. The depth of the plate is denoted by d, giving a rod depth of $d/2$, while the rod width is r. The separation distances are denoted by s_1 and s_2, that measure the gap between the tip of one groove and the bottom of the corresponding groove; s_1 is the separation above, and s_2 below, a plate. Unless otherwise stated, the rods are aligned to the centre of the opposing groove. If they are

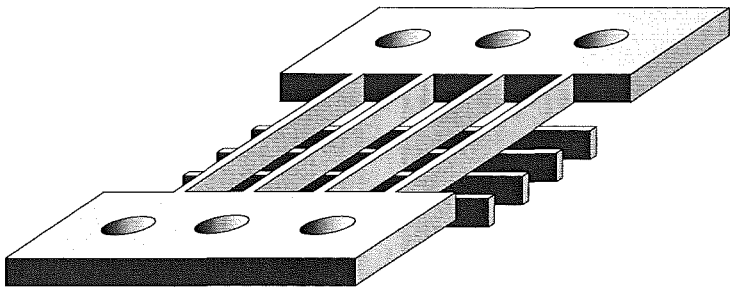


Figure 8.2 A single VPBGF plate has two orthogonal grids. It does not fall apart because the grids are joined. The plate may be actuated by mounting lugs attached to the outside of the plate. A VPBGF is built by rotating each successive plate by 90° so that teeth of adjacent wafers may interlock.

not, the displacement from this position is denoted by l, as shown in Fig. 8.4.

Inside the VPBGF, the volumetric fill factor ξ is given by the dimensions to be

$$\xi = \frac{2rd}{(d + s_1 + s_2)\Lambda}, \tag{8.1}$$

where s_1 and s_2 are the separation distances above and below each layer. Note that l does not affect the fill factor.

In order to successfully measure the behaviour of a VPBGF device in an experimental setup, it is necessary to be able to predict the effect of the plate dimensions on the location of the photonic band gaps and defect modes of interest. This information may be derived in the first instance from simulations.

8.3 SIMULATION TECHNIQUE

8.3.1 Software selection

The previously-used Gsolver rigorous coupled-wave analysis tool was computer-based, and with the available computing resources it proved inadequate for the present photonic crystal modelling task. Algorithmic complexity limitations, such as a quadratic dependence of computation time on the number of diffracted orders retained in the calculation, caused the machine to "hang" when solving structures having several layers of crossed gratings with large period to wavelength ratios (approximately $\sim 2/3 \leq \Lambda/\lambda_n \leq 2$, where the material wavelength $\lambda/n = \lambda_0/n$ where λ_0 is the free-space wavelength and n is the refractive index of the material). Since structures of this size and nature are required to be analysed here, an alternative technique was sought.

Several computer-based numerical codes were available for simulating the VPBGF structures. These included a photonic band structure calculator (MIT Photonic Bands [108]); a derivative of the transfer matrix method (TMM) [111, 190], named Translight; and several finite difference time domain techniques (TEMPEST [191], Remcom XFDTD [192], ISE EMLAB [193]). Note that these are all full vector electromagnetic methods. Since the reflection and transmission coefficients were of interest in this study, the preference was for one of FDTD techniques, although the TMM method would probably have been equally suitable. Of the three FDTD techniques, TEMPEST was the only FDTD code to permit, in the same simulation, the desired combination of periodic and perfectly matched layer (PML) boundary conditions. However, not all of the desired functionality was present in TEMPEST.

8.3.2 TEMPEST FDTD code improvements

The as-distributed version of TEMPEST provided for monochromatic continuous-wave excitation, and steady-state field output. It is time consuming to analyse a broadband

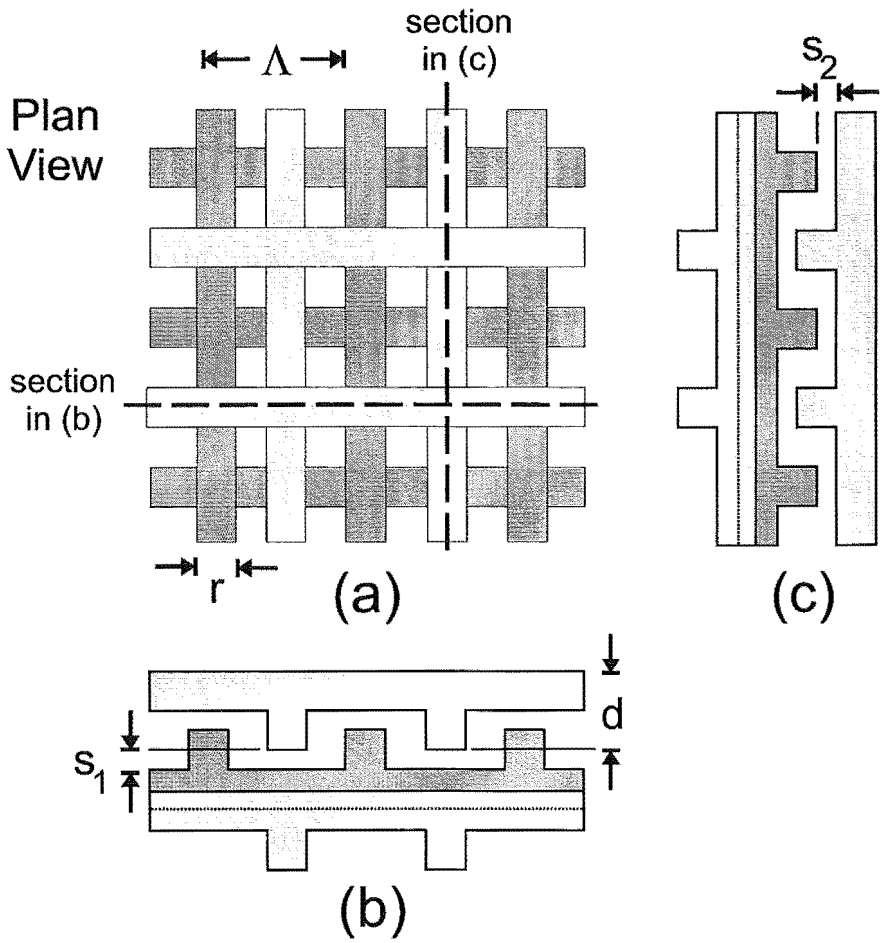


Figure 8.3 VPBGF plate dimensions. Three plates shown in a partially interlocked position, with a plan view in (a) and orthogonal cross-sections in (b) and (c) as indicated. The plate depth is d, while the groove depth is $d/2$. The separation between the interlocking plates is measured from top of tip to bottom of opposite groove. In this case, the separation above (s_1 in (a)) and below (s_2 in (b)) the middle plate are equal ($s = s_1 = s_2$). The period of the rods on a plate is Λ, while r is the rod width.

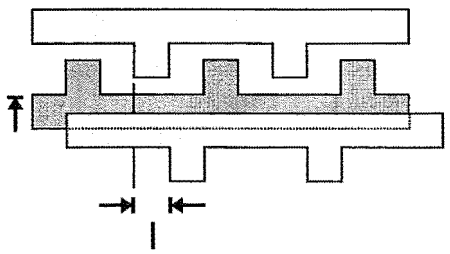


Figure 8.4 The lateral shift between parallel rods in different layers of the VPBGF is given by l. Unless otherwise specified, there is no such lateral shift and $l = 0$.

device, such as a filter, with this "steady-state" method. A faster "transient" method is known, in which broadband excitation is provided by a short pulse, and the electromagnetic field values are sampled at every time step until they have decayed away. Subsequently, the discrete Fourier transform (DFT) may be taken to yield the response of the device across the range of frequencies excited by the pulse [166].

Several new features were added to the TEMPEST source code to enable transient simulations. New keywords, specific to transient simulations, were added to the input file (*.in) handling routine. The control flow was modified so that either steady-state or transient simulations could be run, thereby retaining all of the original functionality. Various new functions were created to manage the progress of the transient simulation, and efficiently write the fields out to disk at each time step. Since the modifications did not affect the core FDTD algorithm, testing requirements were greatly minimised and the modified version was able to be used with confidence. The following example is one of the cases that was used to check the modified code.

8.3.3 Transient simulation example

A simple situation is described in order to explain the FDTD transient simulation technique that will be used for the VPBGF devices. Here, the transmission coefficient of a planar silicon slab is calculated. Ordinarily, a 1-D FDTD simulation domain would be sufficient for this problem, however, the VPBGF requires a 3-D simulation domain and that is tested here.

The 3-D simulation domain is shown in Fig. 8.5, where it is set up to calculate the transmission coefficient of a silicon slab. Boundary conditions (BCs) are specified along each of the six faces of simulation domain, truncating the domain to a computationally practical size. At the top and bottom are absorbing BCs that prevent the outgoing waves from being artificially reflected back into the simulation domain. Berenger's Perfectly Matched Layer (PML) [194] is used here because it matches plane waves of arbitrary incidence, polarisation and frequency. This is important when a potentially-diffractive structure such as the VPBGF is excited by a broad band pulse and a variety of outgoing waves are produced. On the four sides of the domain, periodic boundary conditions are used. This allows a periodic structure such as the VPBGF to modelled more efficiently, since only one unit cell of the device needs to included within the simulation domain. For the silicon slab, there is no underlying periodicity and therefore it simply appears to have infinite extent in the lateral directions (the same assumption is made in the analytical solution).

Whatever device is being modelled, it must be placed so that any material interfaces are spaced approximately one wavelength (or more) away from the nearest PML BC. Otherwise, evanescent waves emanating from the interface may affect the operation of the PML BC and affect the accuracy of the simulation.

Near the start of the simulation, a plane wave wave is launched from the top of the simulation domain. Since a plane wave is produced in only the one direction (towards the silicon slab), and there is no backward travelling wave, the plane wave source may be placed close the PML BC. For a transient simulation, the plane wave has the envelope of a pulse. The subsequent reflected and transmitted pulses are also shown. The electric fields are sampled at the bottom of the simulation domain, and written to file at each time step.

For this simulation domain, the direction of propagation is $-z$, while the x and y directions are parallel to the plane of the silicon slab. Ordinarily, the cell size should be chosen in the range of

$$\lambda_n/10 \leq \Delta x \leq \lambda_n/20, \tag{8.2}$$

where λ_n is the material wavelength in the most optically dense region of the simulation domain, and $\Delta x = \Delta y = \Delta z$ are the x,y and z dimensions of a single, cubic, FDTD cell. For the undesirable case of coarse spatial sampling, $\Delta x > \lambda_n/10$, the wave undergoes nonphysical exponential decay while propagating and the numerical phase velocity may exceed c, both leading to unreliable results. If it is desired to analyse devices containing materials no more optically dense than silicon ($n_s = 3.42$), at frequencies up to the W-band (110 GHz), then maximum acceptable cell size is $\Delta_x = 79.9$ μm. Here, a cell size of $\Delta_x = 37.5$ μm was chosen to meet this restriction, and provide convenient cell sizes for modelling VPBGF devices that it was intended to construct from 525 μm thick silicon [1].

The Courant stability criteria [195]was met by choosing a time step of 7.2212×10^{-14}s. This gave a Courant stability number of 0.6, well below the critical value of 1.0, above which FDTD simulations become unstable.

8.3.4 Validation results

Two simulations were run in order to obtain the transmission coefficient of the silicon slab, one without the slab and one with. The same illumination was used in each case (a Gaussian pulse plane wave), and the transmitted electric fields were recorded for a "real time" of 500 ns, or 6924 time steps. The intensity of the transmitted electric fields are plotted in Fig. 8.6. The solid line is for the empty (air-filled) domain, while the dotted line is for the domain with the silicon slab.

From the first principles of plane wave propagation, it may be seen that these pulses are qualitatively as expected. For example, the Gaussian pulse is launched 60 ps after the start of the simulation, and must then traverse 205 cells of FDTD space, equivalent to 7.69 mm, before it is detected by the field sampling routine. Assuming

[1] The silicon fabrication task is part of on-going work and reported upon in this thesis

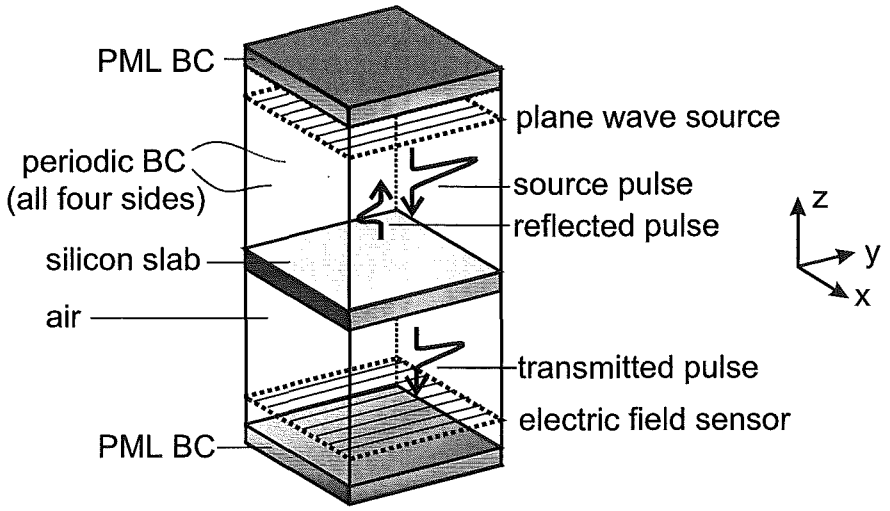


Figure 8.5 Visualisation of the FDTD simulation domain for the analysis of a slab of silicon.

a propagation speed of c, this is expected to take 25.6 ps. It may be seen that in the empty simulation, the pulse is detected at 85.6 ps after the start. In the silicon slab simulation, the first of several pulses is detected 1.7 ps later, at 87.3 ps, because of the additional delay imposed by propagating through the 225 μm-thick silicon slab. Subsequent pulses arising from multiple reflections within the slab are delayed by a further 5.1 ps in each case, this being the time required to traverse the silicon slab twice.

The discrete Fourier transform (DFT) of the pulses may be taken in order to yield the frequency response in Fig. 8.7. Details of the DFT algorithm may be found elsewhere [196]. The frequency resolution of the DFT is determined by the step size, and the number of time steps, according to the following relation

$$\Delta f = \frac{1}{N\Delta t}, \tag{8.3}$$

where N is the number of time steps input to the DFT algorithm. Usually, the fast Fourier transform (FFT) is used because it is faster than the DFT, although the restriction is that the algorithm only works for sequences that have a length that is a power of two, i.e. $N = 2^p$, where p is an integer. Therefore, the 6924 steps produced by the simulation are zero-padded until the length reaches $N = 8192 = 2^{13}$. Thus, Δf = 1.6904 GHz for the present simulation.

It may be seen in Fig. 8.7 that the frequency response of the silicon slab is weighted by the power spectral density of the incident wave, and therefore does not give the transmission coefficient directly. However, the two traces may simply be divided to

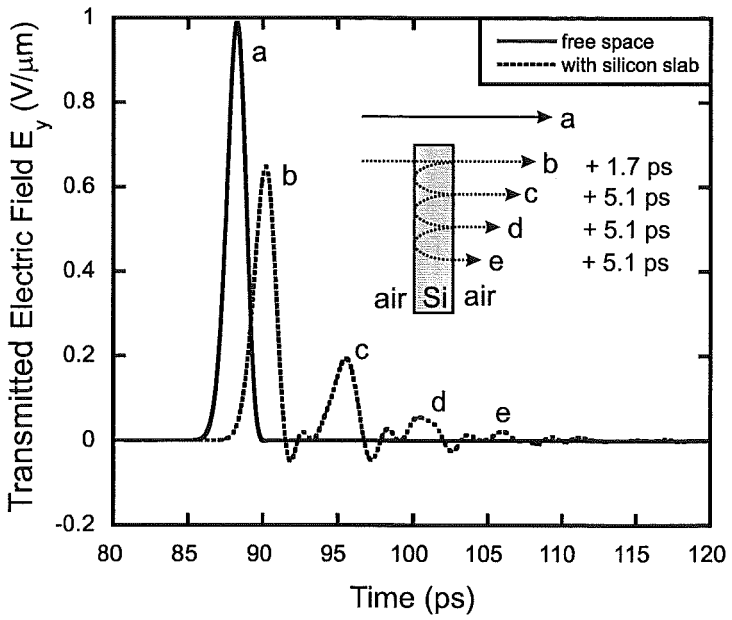


Figure 8.6 The recorded electric field amplitudes of a Gaussian pulse propagating in free space (a) and through a finite thickness silicon slab (b - e).

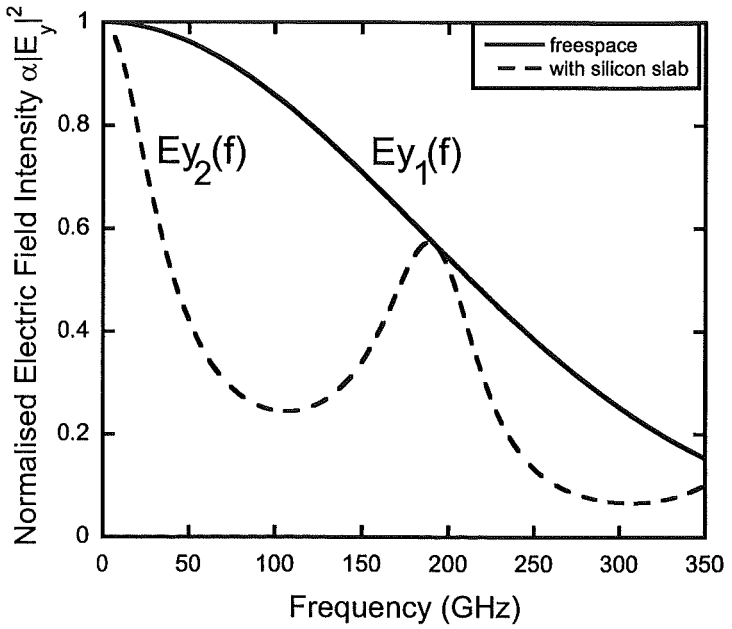


Figure 8.7 The normalised electric field intensity as a function of frequency for a Gaussian pulse propagating in free space, and through a finite-thickness silicon slab.

give the desired transmission intensity coefficient $T(f)$ as follows

$$T(f) = E_{y1}(f)/E_{y2}(f), \tag{8.4}$$

where $E_{y1}(f)$ is the transmission intensity of the free space pulse, and $E_{y2}(f)$ is the transmission intensity of the pulse after transmission through the silicon slab. These results are presented in Fig. 8.8.

8.3.5 FDTD error estimation

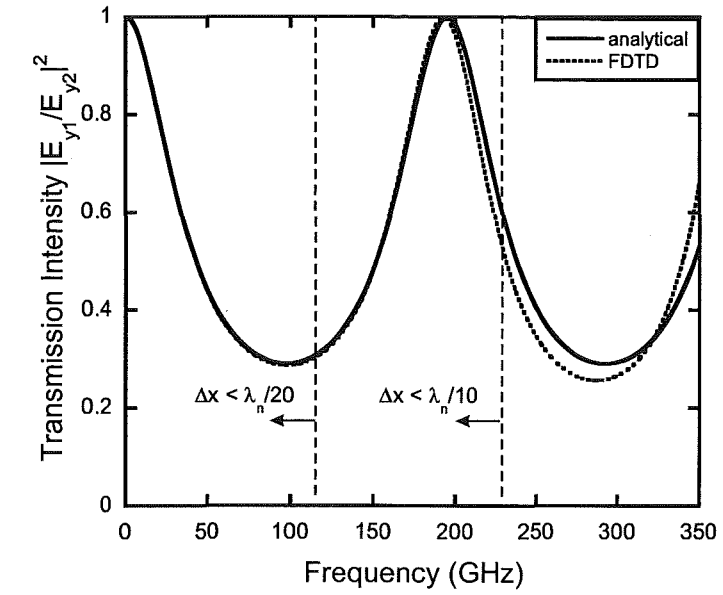
The error of the transient simulation technique may be estimated by comparison with the exact results given by an analytical or semi-analytical method, such as the T-matrix method of section 6.2.1. The results are plotted in Fig. 8.8, with the FDTD and analytical transmission coefficients in Fig. 8.8(a), and a magnified view of the error in Fig. 8.8(b). The relationship between the fixed cell size (37.5μm), and the broad range of wavelengths excited by the pulse, is indicated by vertical bars indicating the thresholds of $\Delta x \leq \lambda_n/10$ and $\Delta x \leq \lambda_n/20$. The error is less than 1.5% until approximately 140 GHz, and is no greater than 2.2% up to 200 GHz. Therefore, this particular simulation setup could be expected to give results of sufficient accuracy for design purposes up to approximately 200 GHz, but with more accurate results suitable for comparison with experiment at the W-band measurement frequencies (70 – 110 GHz).

8.3.6 Extension to photonic crystals

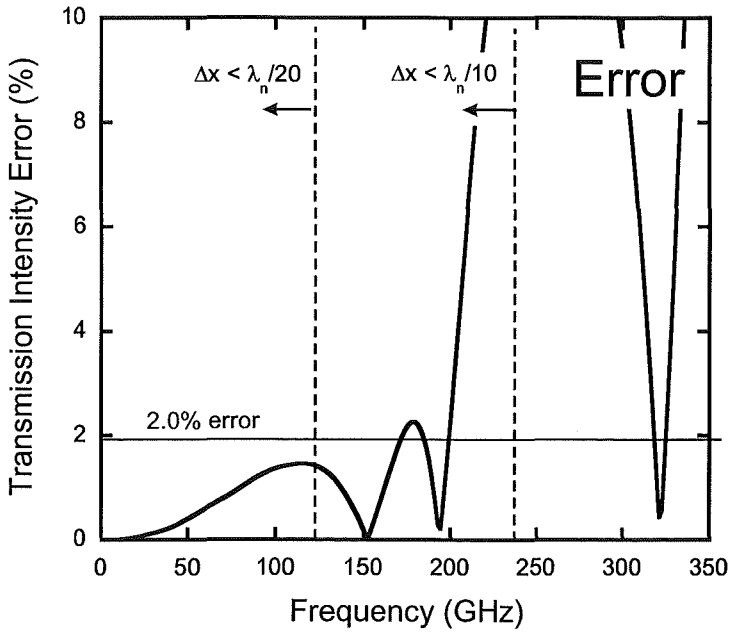
Now that an estimate of the confidence levels in the transient simulation technique has been developed, it is possible to turn to the analysis of photonic crystals. This section shows that the TEMPEST FDTD technique may be used to successfully model photonic crystals, by presenting the results of a simulation of a face-centred-tetragonal (fct) woodpile of the type proposed by Ho *et al.* [114].

The fct woodpile simulated here, shown as an inset in Fig. 8.9, comprises three unit cells of four layers of rods each, for a total of twelve layers of rods. Each layer is 262.5 μm deep, and the rods are 400 μm wide, and separated by an air gap of 800 μm to give a period of 1.2 mm. The direction of propagation is normal to the stacking the direction.

Figure 8.9 shows the simulated transmission coefficient of the woodpile stack for frequencies 0 – 200 GHz, the range for which the simulations are sufficiently accurate for design purposes. For photonic crystals, the presence of a photonic band gap is indicated by a band of frequencies for which the transmission intensity is small ($T < 0.01$). For this woodpile, the lower edge of the photonic band gap may be seen at 110 GHz, while the characteristic Fabry-Pérot ripple is evident frequencies below the band gap. Plotted



(a)



(b)

Figure 8.8 Comparison of FDTD and analytical predictions of the transmission coefficient for a $225mum$ thick silicon slab ($n_s = 3.42$). (a) The FDTD and analytical transmission coefficients are plotted along with an indication of the apparent spatial sampling rate in the FDTD calculation. (b) The error between the exact analytical solution and the FDTD solution.

on a log scale (not shown), the mid gap rejection rate is 51dB, which is consistent with the accepted -17 dB attenuation per unit cell [116]. These results provide additional validation of the FDTD code and further confidence in the results for the VPBGF structures.

8.4 VPBGF PLATE DESIGN

During the VPBGF plate design process, it is often necessary to adjust the dimensions (such as rod depth, width, period, refractive index, and number of layers thereof) so that the tunable features of the eventual device may be located within an experimentally observable frequency range, such as the W-band (70 – 110 GHz). In this process, it is helpful to know the effect of increasing or decreasing any of the plate dimensions ahead of time. For example, it is known that for structures such as the fct woodpile of Ho *et al.*, increasing the stacking periodicity will lower the band gap frequency [114]. Similar relationships are expected to apply to the VPBGF, and are found here to do so.

The effect of varying each of the fixed plate dimensions is investigated, and the results are presented in band gap map format. Band gap maps are a concise method of showing how a the variation of parameter affects the frequency and width of a photonic band gap. A series of simulations is performed, with a single parameter being varied across the simulations. Regions of low transmission ($T < 20$ dB) are identified as belonging to a photonic band gap.

All of the band gap maps in this section are plotted for a typical 6-plate (12-layer) VPBGF. Unless otherwise stated, the device has a substrate refractive index of $n_s = 3.42$ (the other medium is air, $n_i = 1$), a period of Λ=1.2mm, plate depth of $d = 525\mu$m, rod width of $r = \Lambda/4 = 400\mu$m, and everywhere an equal separation between the wafers of $d/2 = 262.5\mu$m. Since the device dimensions scale linearly with wavelength, and silicon exhibits minimal dispersion in the frequency range of interest, the frequency given is dimensionless, i.e. it does not have an absolute value in Hz, but is related to the dimensions of the structure, usually via Λ or d.

8.4.1 Substrate refractive index

The effect of increasing the refractive index contrast is illustrated in Fig. 8.10, where the plates are constructed from a dielectric having $n_s > 1$, and the other medium is air $n_i = 1$. The method of band gap map construction is illustrated in Fig. 8.10(a). The error in the location of the calculated photonic band gap edges in Fig. 8.10 (and following maps) is estimated to be ± 3 dB and ± 2 GHz, which is acceptable for the purposes of illustrating behavioural trends of these devices.

For index contrast ratios below $n_s/n_i \sim 1.9$, there is no observable gap; this is in line with existing observations for 3-D photonic crystals [114]. The TM band gap

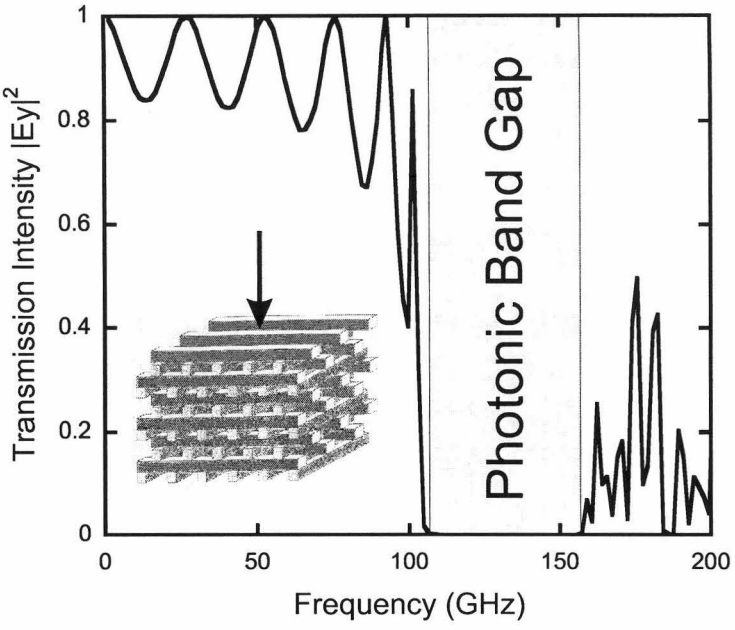
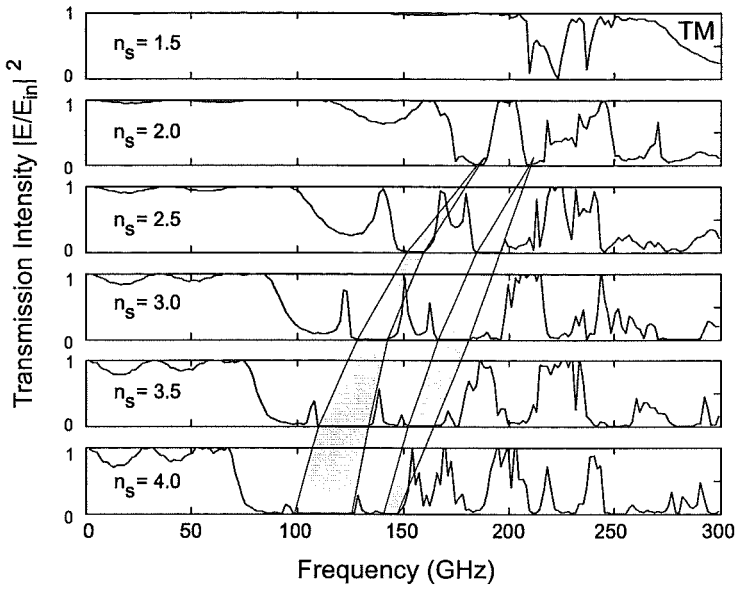


Figure 8.9 Transmission intensity of a 12-layer silicon-air fct woodpile photonic crystal (pictured, inset), calculated using TEMPEST Finite Difference Time Domain (FDTD).

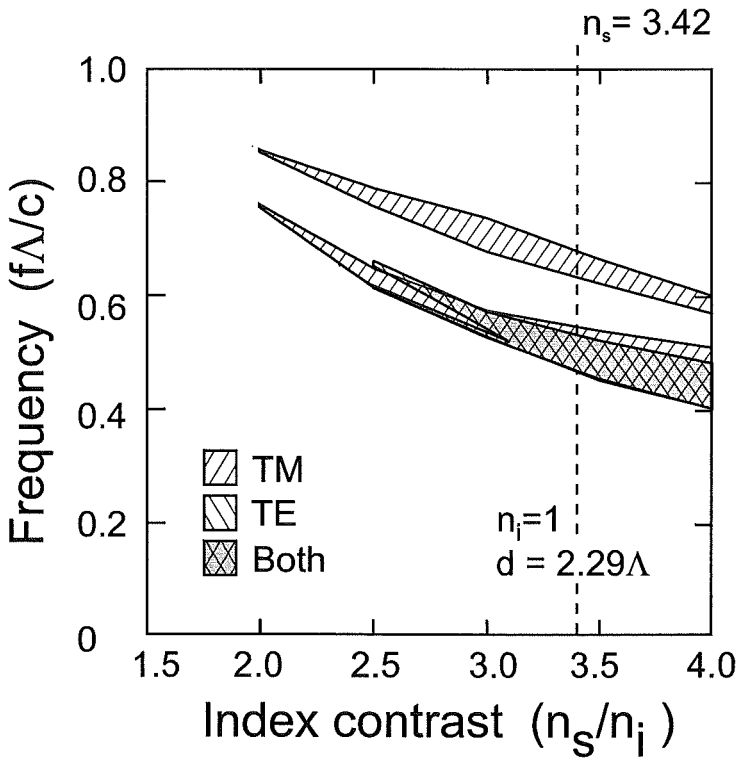
does not open up until the index contrast reaches 2.5. It may be seen that there is a polarisation dependence that would be unexpected if the structure was infinitely thick. However, it has finite thickness, and the direction of the rods on the terminating faces of the crystal have an effect on the properties. Therefore, the remainder of the results in this section will include both polarisations. Overall, increasing the index contrast is seen to draw the band gap lower in frequency. This is attributed to the effective size of the structure appearing larger as the material wavelength shortens. It is reported in the literature that increasing the size of the unit cell of a woodpile reduces the band gap frequency [114], therefore this behaviour is not unexpected. The influence of the refractive index may be exploited to create a tunable device, by flooding a structure (constructed from a material having fixed-refractive index) with a fluid having a variable refractive index. For example, the refractive index of "liquid crystals" may be controlled by the application of an electric field. This method has already been reported in conjunction with photonic crystals [131]

Returning to the VPBGF, the main point of this section is that a substrate with a refractive index of at least 1.9 must be chosen (assuming that the other medium is air, $n_i = 1$). Otherwise no band gap will form. Aside from that, there is great flexibility in the choice of material, and other criteria such as low loss and low dispersion become important. For example, silicon ($n_s = 3.42$) is practically lossless and has low dispersion at terahertz frequencies.

For a silicon-air structure ($n_s/n_i = 3.42$) having the dimensions in the caption to Fig. 8.10, the band gap is centred at 125 GHz, which is outside of the W-band range



(a)



(b)

Figure 8.10 Band gap maps showing the effect of varying the refractive index contrast. The device simulated here has six plates (12 layers of rods). The rods have period $\Lambda = 1200$ μm, width $r = 400$ μm, and depth $d/2 = 262.5$ μm. The plates are equally separated by $s_1 = s_2 = d/2$. The substrate refractive index is varied in the range $1.5 < n_s < 4.0$ while keeping constant the refractive index of the other medium (air, $n_i = 1$). (a) The simulations from which the band gap map is constructed (TM only). A filled polygon is drawn between the points that represent the upper and lower edges of the photonic band gap; (b) The resulting band gap map (rotated by 90° with respect to (a)). The refractive index contrast of a silicon-air device is indicated by the dashed vertical line.

(70 – 110 GHz) in which it is desired to measure a prototype. This situation could be marginally improved to 112 GHz (the lower band edge would be within the W-band) by finding a material with a higher refractive index of 4.0, however this is not a trivial exercise. Instead it is preferable to stay with the choice of a silicon substrate, with all of the advantages thereof (low loss, low dispersion, micromachinable), and investigate how the device dimensions, such as the width of the rods, may be adjusted in order to bring the band gap into the W-band measurement range.

8.4.2 Rod width

The effect of increasing the rod width r (and hence fill factor ξ) is shown in Fig. 8.11. As the rod size increases, the centre frequency of the band gap is drawn down in frequency. The effect is most noticeable for small rod widths ($r < 0.1\Lambda$). Since there can be no band gap if there are no rods, the width of the band gap reduces as the rod width approaches zero. The exact point at which the band gap is extinguished is not identified by the present simulations due to the finite cell size of the FDTD lattice. Rod widths greater than $\Lambda/2$ are not investigated because these structures cannot interlock, and it is desired here to investigate tuning methods that arise when the plates are able to interlock.

A typical rod width for a woodpile is 25% of the rod period, or $r = \Lambda/4$. The centre frequency of the band gap for the simulated device (see caption of Fig. 8.11 for dimensions) is 125 GHz, with a TE band width of 25 GHz. For this device If the rod width was halved to $r = \Lambda/8$, then the centre frequency would rise to 150 GHz. Thus the rod width may be adjusted to place the centre of the band gap anywhere in the F-band[2] (120 – 150 GHz), yet not require overly thick rods. However, it is desired to measure the device in the W-band (70 – 110 GHz), so it appears necessary to increase the rod width.

At the maximum thickness of 50% of the period, the centre frequency of the band gap is 100 GHz, which falls into the W-band range as desired. However, there are some problems with using such thick rods. Firstly, if the substrate is at all lossy, then having such a high concentration of dielectric material in the device will cause it to have a higher insertion loss than an equivalent device with thinner rods. Note that the effect of loss in the substrate is not explored here, since it is possible to obtain high-resistivity ($R > 5000$ Ω-cm) silicon that is practically lossless [16]. Secondly, with such thick rods it would no longer be possible to shift the plates laterally, and this would eliminate a potential tuning method. Therefore, it is necessary to investigate whether a typical rod thickness of $r = \Lambda/4$ may be maintained while other plate dimensions are adjusted to bring the centre frequency of band gap into the W-band range.

[2]The F-band designation is used elsewhere for alternative frequency ranges, and should not be confused.

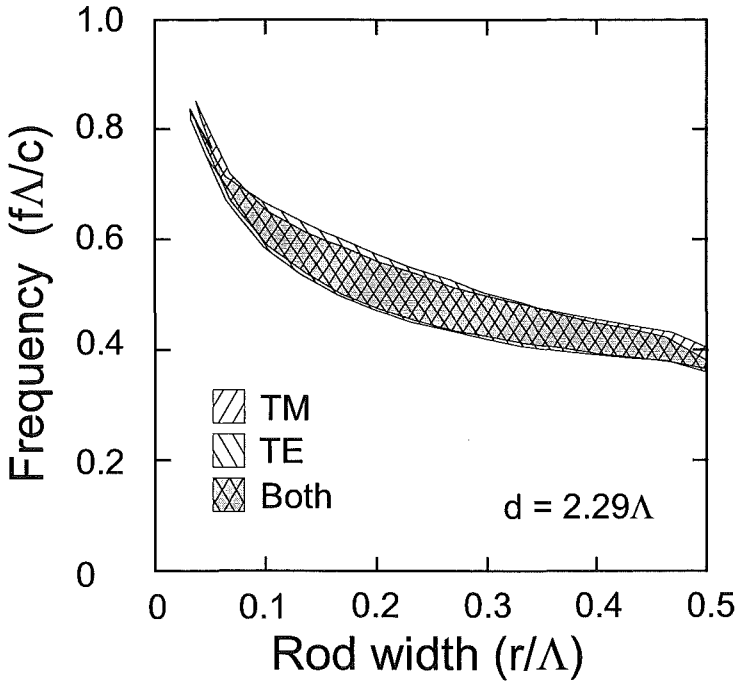


Figure 8.11 Band gap map showing the effect of varying the rod width. The device simulated here has six plates (12 layers of rods). The rods have period $\Lambda = 1200$ μm and depth $d/2 = 262.5$ μm, but the width is varied from $0 < r/\Lambda < 0.5$. The plates are equally separated by $s_1 = s_2 = d/2$. The substrate refractive index is silicon ($n_s = 3.42$) while the other medium is air ($n_i = 1$).

8.4.3 Rod depth and period

Increases in either the rod depth or period also cause the band gap to drop in frequency. The band gap map in Fig. 8.12 shows the drop due to increasing the rod depth, while holding the period constant. This is consistent with the literature [114]. It should be emphasised that while a rod depth to period ratio is plotted on the independent axis, the use of the ratio is not intended to imply that an increase in the period raises the actual band gap frequency. In fact, by recasting the same data in terms of the reciprocal ratio, it may be seen in Fig. 8.13 that an increase in the period (relative to the depth) also causes the band gap frequency to drop. Note that the dimensionless frequency F is defined differently in Figs. 8.12 and 8.13 (see the captions).

These relationships are useful to know, because it suggests a flexibility in fixing fabrication parameters. For example, if given a fixed wafer thickness, e.g. due to availability, the period may be adjusted place the band gap within a frequency range for which measurement equipment is available. For example, given 100mm diameter silicon wafers having a (standard) thickness of 525 μm, then without thinning the wafer, the rod depth would be fixed at 262.5 μm. Specifying a rod width equal to 0.25Λ, the band gap of this six-plate device could be placed in the W-band (70 – 110 GHz) setting the period to approximately 1.5mm – 1.8 mm, or in the F-band (120 – 150 GHz) by choosing a period of approximately 1.4mm – 1.2 mm. If the period was even further

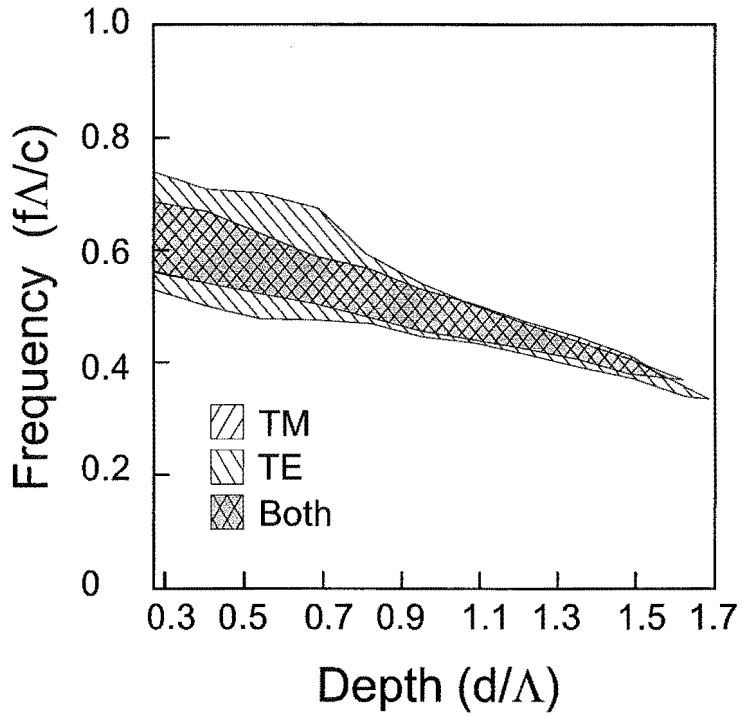


Figure 8.12 Band gap map showing the effect of varying the rod depth to period ratio (d/Λ)). The device simulated here has six plates (12 layers of rods). The rods have period $\Lambda = 1200$ μm and width $r = 400$ μm, but the plate depth (twice the rod depth) varies over the range $0.3 < d/\Lambda < 1.7$. The plates are equally separated by a half rod depth $s_1 = s_2 = d/4$. The substrate refractive index is silicon ($n_s = 3.42$) while the other medium is air ($n_i = 1$). The dimensionless frequency F is defined with respect to the rod period Λ.

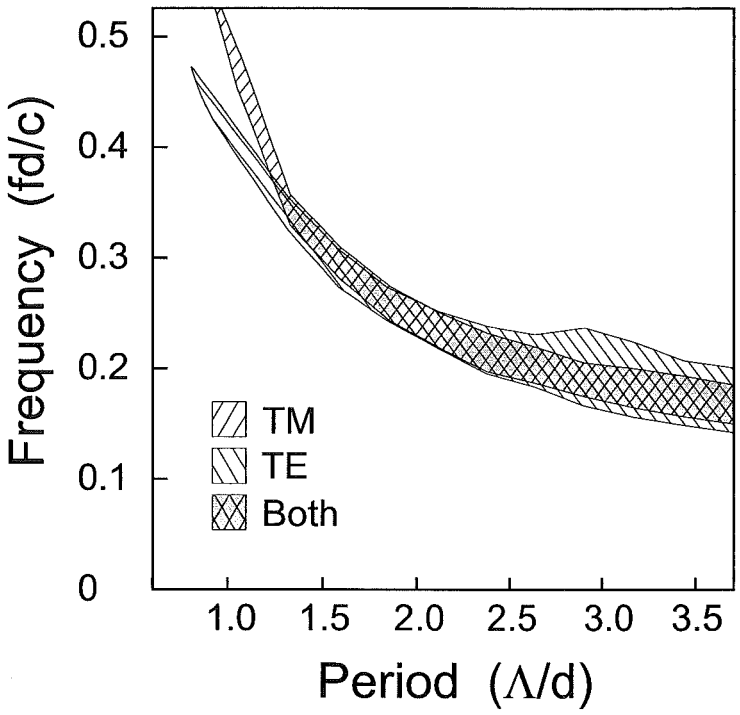


Figure 8.13 Band gap map showing the effect of varying the rod period to depth ratio. The device simulated here has six plates (12 layers of rods). The rods have period $\Lambda = 1200$ μm and and width $r = 400$ μm, but the plate depth (twice the rod depth) is varied over the range $0.5 < \Lambda/d < 3.7$. The plates are equally separated by $s_1 = s_2 = d/2$. The substrate refractive index is silicon ($n_s = 3.42$) while the other medium is air ($n_i = 1$). The dimensionless frequency F is defined with respect to the rod depth d.

reduced ($\Lambda = 500$ μm), the band gap could be placed as high as 250 GHz.

8.4.4 Number of plates

The effect of increasing the number of layers is shown in the discrete-valued band gap of Fig. 8.14. The effect of increasing the number of plates is to produce photonic band gaps at lower frequencies. For the simulated structure (see caption to Fig. 8.14 for dimensions), photonic band gaps may be achieved with as few as just 3 plates (6 layers of rods) at 112 GHz (dimensionless frequency $F = 0.45$), but would require 12 or more at 30 GHz ($F = 0.12$). Since the band gap frequency is referenced to the period, it may be seen that a device with a larger period will require fewer plates to exhibit a band gap at a particular frequency. This is useful because it allows a compact device having less plates to be produced by virtue of having a large rod period. In order to shift the frequency of the 6-plate device down slightly from 112 GHz so that it was in the W-band (70 – 110 GHz), then the period could be increased from the current 1.2 mm to in the range 1.5 – 1.8 mm, as described in section 8.4.3.

8.5 VPBGF ACTUATION SCHEMES

In order to tune the VPBGF "on-the-fly", the plate separation and lateral position may be varied during operation. There are a great many degrees of freedom in the actuation of a VPBGF, and it is not possible to cover them all in this initial investigation. However, an understanding of the capabilities of the VPBGF concept may be derived from investigating the basic schemes presented here.

Before proceeding, it is useful to illustrate at least one possible method for achieving control over plate separation in a real device. For example, plates constructed with integral mounting lugs (such as in Fig. 8.2) could be "threaded" onto a set of rods, as shown in Fig. 8.15. Some of the rods would act as guide rails for alignment, whilst others would be attached to the lugs and actuated in order to adjust the stack.

Using such an actuation scheme, the plate separations could be arranged according to a number of schemes. Four basic tuning schemes (all novel) are presented in Fig. 8.16. In order to emphasise the movements, the details of the gratings are omitted. The initial position of the device is shown in Fig. 8.16(a), where the plates are evenly spaced by $s = d/4$ and partially interlocked. Expanding and contracting the stack, whilst maintaining equal plate separations, gives the "unison" mode in Fig. 8.16(b). Shifting only "every second plate" gives the unequal spacings seen in Fig. 8.16(c). Fixing the separation between all plates, except for two in the middle of the stack, gives the "two groups" mode of Fig. 8.16(d). Adjusting not the separation distance, but the lateral alignment, gives the "lateral shift" mode of Fig. 8.16(e). A key to the

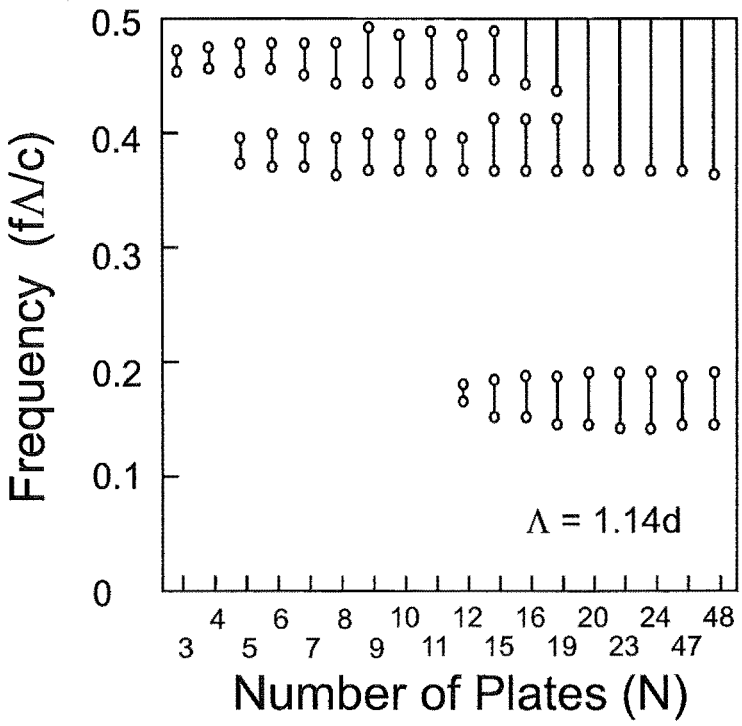


Figure 8.14 Band gap map showing the effect of increasing the number of plates in the VPBGF. The devices simulated here have between 3 and 48 plates (6 and 96 layers of rods). The rods have period $\Lambda = 1200$ μm, width $r = 400$ μm, depth $d/2 = 262.5$ μm. The plates are equally separated by a half rod depth $s_1 = s_2 = d/4$. The substrate refractive index is silicon ($n_s = 3.42$) while the other medium is air ($n_i = 1$). The dimensionless frequency F is defined with respect to the rod period Λ.

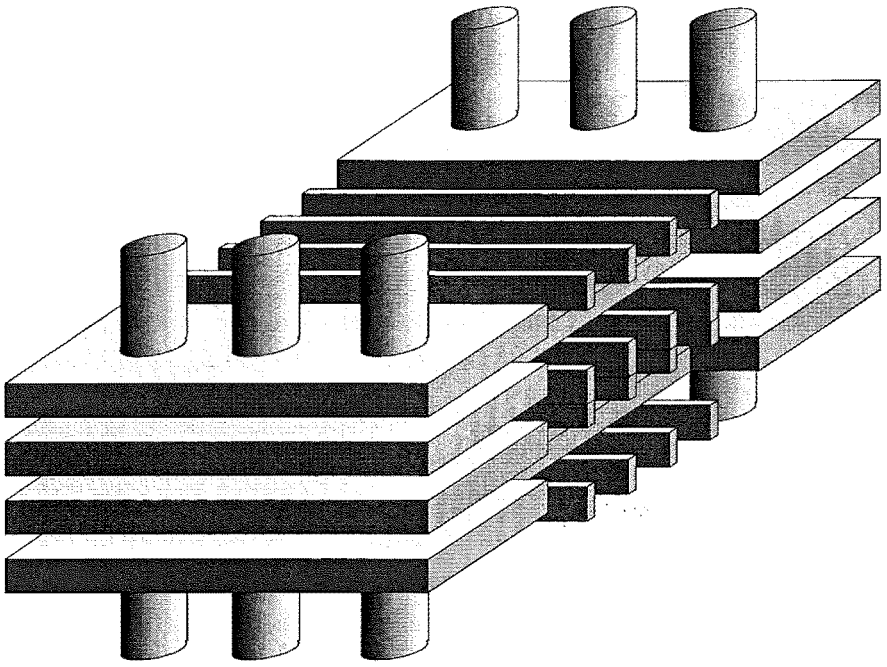


Figure 8.15 The VPBGF device could be actuated by threading the plates onto a series of rods, some of which will move the plates, others which will simply keep the stack aligned.

symbols is provided in Fig. 8.16(f). Some basic properties of device tuning using these different actuation methods are presented in the rest of this section.

8.5.1 Unison

The unison mode of operation allows the band gap to be shifted up and down in frequency. This is achieved by expanding and contracting the VPBGF in the manner of an accordion, all the while ensuring that the separation between plates is uniform throughout the device. This may be illustrated by the three unit cells that have been drawn in Fig. 8.17. The direction of propagation is normal to the stacking direction. The grating vector $\mathbf{K}$ is defined with respect to the grating on the top face of the device. The period of the rods is Λ. The unit cell corresponds to a stack of a half plate, a full plate, and another half plate, to give two plates ($N = 2$) in total, or four layers of rods. Typically a device would be at several unit cells thick, and full plates would be used at the top and bottom of the stack, rather than half plates suggested by the bulk unit cell.

In Fig. 8.17(a) the separation is $s = 0$, giving an overall height of d. At this separation, the VPBGF appears as a simple st woodpile with a period equal to half that of the rods, $\Lambda/2$, and the volumetric fill factor ξ is at its highest, equal to $2r/\Lambda$, or twice the rod width to period ratio. In this position the structure has the highest effective refractive index of all the separations.

As the plate separation is increased, the micro cavities of air in the structure grow in size. The rods remain partially interlocked until the separation becomes equal to the rod depth, at $s = d/2$ as shown in Fig. 8.17(b). At this separation, the volumetric fill factor ξ is the half of the previous value, and now equal to the rod width to period ratio. The overall height of the unit cell has doubled, to $2d$.

The rods do not necessarily need to remain interlocked, and once the separation depth exceeds the rod depth, it is said that the plates are "separated". An example is shown in Fig. 8.17(c) for the case of a separation distance equal to the plate thickness, $s = d$. The separation distance has been doubled, but the overall thickness only increases by 50% and ξ falls to two thirds of the rod width-period ratio. Thus the rate of change of volumetric fill factor is greatest for the separation distances less than the rod depth.

The operation of a unison-mode device is summarised in the band gap map of Fig. 8.18. The particular device is constructed with six plates, having a period-depth ratio of 2.29. It may be seen that the device is sensitive to the polarisation, since the band gap in the TE polarisation opens up at smaller separations than for the TM polarisation. This is attributed to having an even number of layers in the stack which causes the exposed layers of rods on the top and bottom surface to have the same orientation, preventing the 90° rotational symmetry required for TE and TM

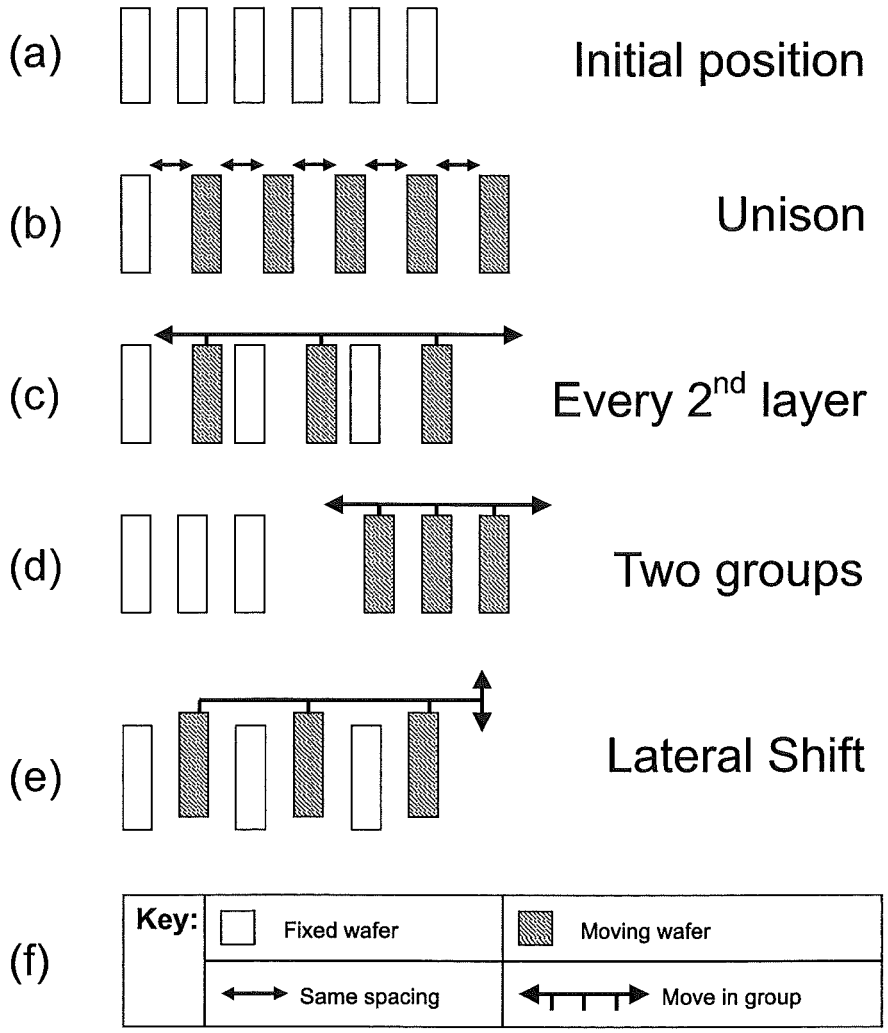


Figure 8.16 Four schemes for tuning the VPBGF "on-the-fly". For clarity, details of the gratings are omitted, but it is understood that the gratings are, for the most part, interlocking. The initial position of the stack is given in (a), while the four schemes are given in (b) – (e). A key to the symbols is given in (f).

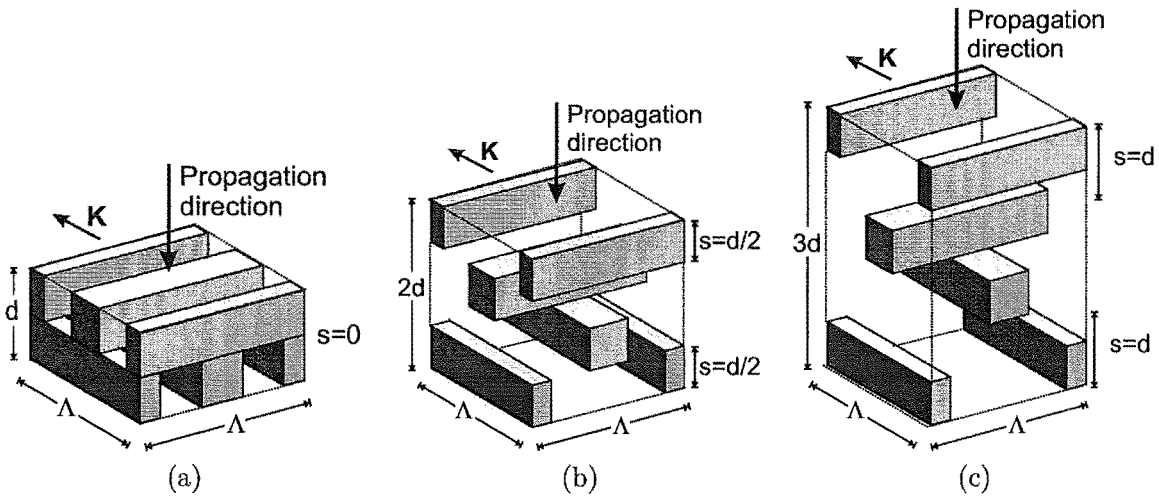


Figure 8.17 Three representative unit cells for the "unison" mode of actuation. The spacing between all adjacent layers is always the same throughout out the device, so as the separation is varied, the height of the unit cell varies too. (a) $s = 0$, (b) $s = d/2$, (c) $s = d$.

polarisation invariance. The sensitivity is expected to reduce if an odd number, or a larger number, of plates is used, although these are not explored here.

For the interlocked plates, there are two different band gaps, one just above $F = 0.2$, and one just below $F = 0.3$. In the separation region, the two bands merge at separations $s/d > 0.7$. This is a single band that has been split in two by a broad defect mode transmission resonance. The broadness of the defect is consistent with there being many layers of defects. For separations less then the rod depth, the defect arises from the interlocking rods, while for separations greater than the rod depth, the defect arises from the extra air gaps between layers.

While the upper band edge is relatively constant, the three other band edges may be seen to decrease in frequency by about $0.076F$ over the range of separations studied. For W-band device with $d = 525$ μm, this would correspond to a shift of about 4 GHz, from 114 – 110 GHz. The band edge shifts down in frequency because the stacking periodicity of the structure has increased.

Unfortunately, the unison mode does not produce a great deal of tuneability, and is further disadvantaged by requiring N-1 separate actuators, where N is the number of plates. This would result in a complicated, and therefore expensive, mechanical design.

8.5.2 Every-second-plate

In the "every second plate" mode of operation, only one actuator is required and the width of band gap may be considerably altered. Since only every second plate is moved, the device thickness varies a great deal less than the unison mode. Three representative unit cells are shown in Fig. 8.19, where it may be seen that a higher-

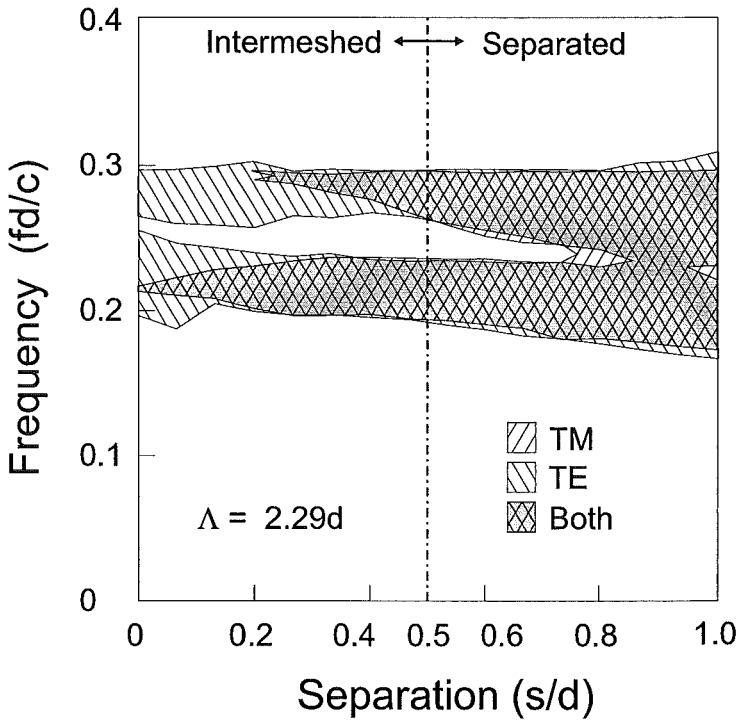


Figure 8.18 The operation of a unison-mode VPBGF. The defect layers created by the variable separation give a defect mode that splits the band for separations up to $s/d = 0.7$. The uppermost band edge is relatively constant in frequency, but the other edges show a downward shift as the separation increases.

than-average refractive index defect layer is created at the bottom of the unit cell where the rods interlock, while a lower-than-average refractive index defect layer is created at the top of the cell in the air gap. Thus, both higher-than-average and lower-than-average refractive index defect layers are present together in both the settings shown in Figs. 8.19(a,b). In Fig. 8.19(c), there is no defect layer.

The band gap map for the "every second plate" mode of operation is plotted in Fig.8.20. For $s/d = 0.5$, the structure and band gap are the same as for unison mode. Again, there is a single band split by a broad defect mode transmission resonance. However, for this device, there are important differences in the performance. For the TE polarisation, the defect mode is clearly associated with the upper band edge, and is therefore caused by a defect that adds (donates) substrate material. This is then the mode that is associated with the overlapping rods at the bottom of the unit cell, since this is the densest point in the structure. With increasing separation, this overlapping layer gets thinner, and the defect mode moves further towards the middle of the band gap, becoming a "deep" donor defect.

For the TM polarisation, the defect mode is more closely associated with the lower band edge. This then is an acceptor, associated with the air defects. As the separation increases, the air layer also becomes thinner, and the defect mode moves to the centre of the band gap and shrinks in width. Since the TM band gap disappears below $s = 0.1d$, this device could be used to modulate the TM polarisation, i.e. switch it "on" and "off". For filtering, it would be more useful to have a tunable narrow passband, but unfortunately, the defect modes produced in the unison and every-second-plate modes are too broad for this purpose.

8.5.3 Two-groups

The width of a defect mode resonance may be narrowed by decreasing the number of similar defects in the structure. In the "two groups" mode of operation shown in Fig. 8.21, the wafers are actuated such that only the teeth at the interface between the two groups interlock, and therefore there is only one layer of defects. In Fig. 8.21(a) a maximally-thick defect layer is created by fully interlocking the rods on the faces of the two groups. The two groups are shaded differently from each other to aid identification. In Fig. 8.21(b), a slightly thinner defect layer is shown, for the case where the separation between the two groups is $s = d/4$.

Such a defect mode has not been studied in the literature before, although the related planar air cavity defect has been investigated by a number of authors [186,197]. The present work is an extension to these studies, but the nature of the defect is fundamentally different. Within the planar air cavity, the index of refraction is lower than the average value for the crystal, whereas for the VPBGF two-groups defect, the defect has a higher effective index of refraction than the rest of the crystal and the

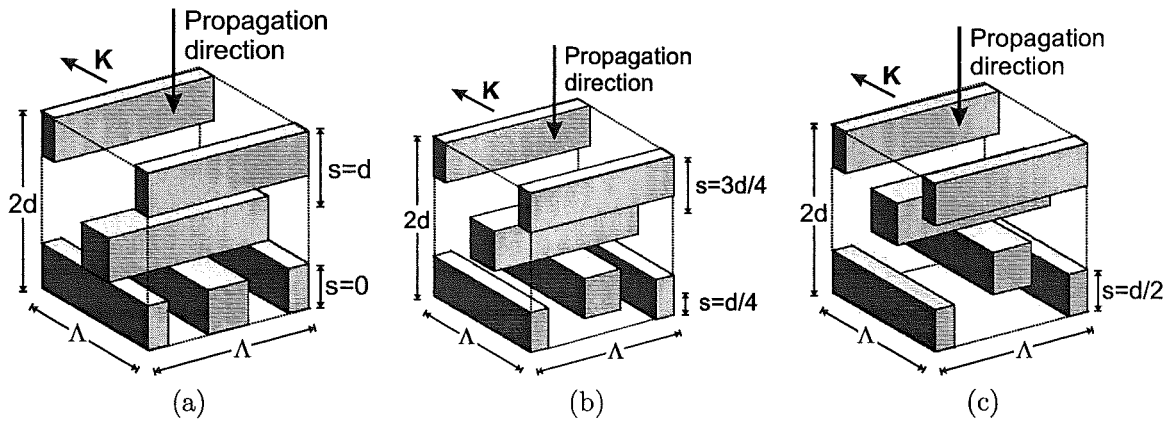


Figure 8.19 Three representative unit cells for the "every second plate" mode of actuation. The separation distances on either side of each layer always add to a constant value, in this case d. The separation distance at the bottom of the unit cells are as follows (a) $s = 0$, (b) $s = d/4$, (c) $s = 3d/4$.

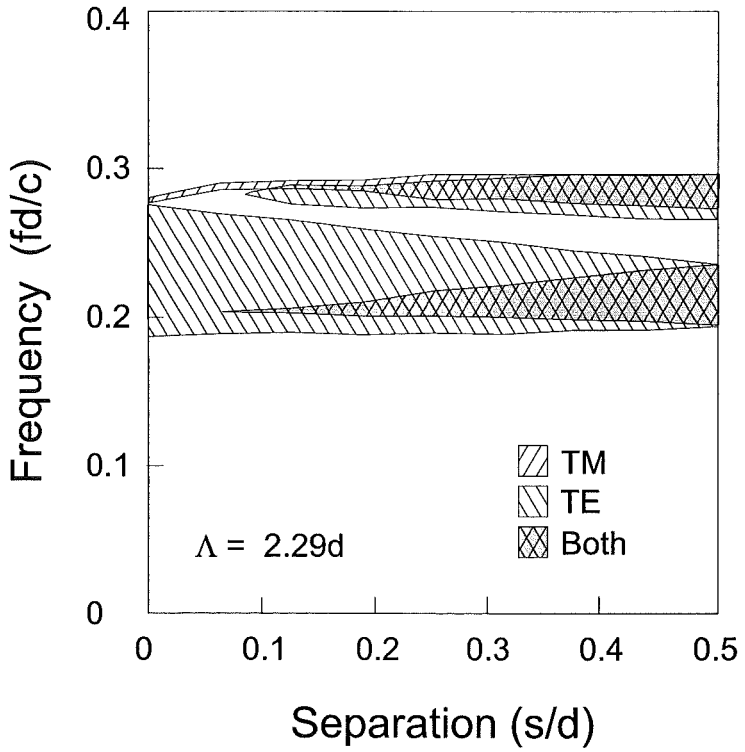


Figure 8.20 The operation of the VPBGF in "every second plate" mode. The TE and TM bands are both split by a defect mode. The net effect is to give a variable bandwidth in both TE and TM, but of the opposite sense to each other.

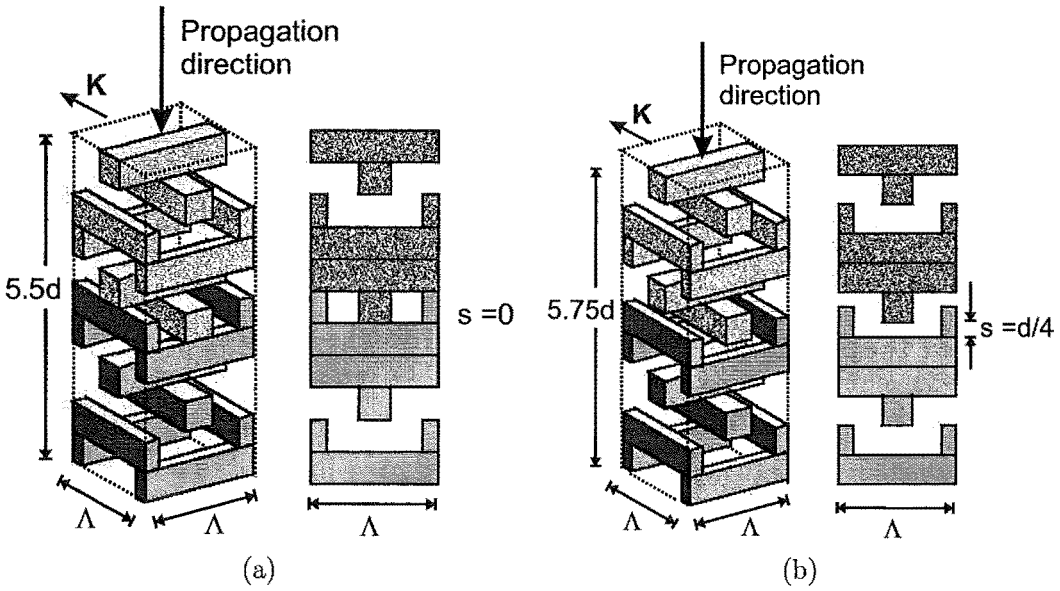


Figure 8.21 Unit cells of a six-plate stack actuated in the "two groups mode" to produce a central defect layer with a higher-than-average refractive index. (a) $s = 0$ for maximum-thickness defect layer, (b) $s = d/4$ thinner defect layer.

defect periodicity is twice that of the supporting photonic crystal. Such double-period higher-than-average refractive index defects appear to be of use in applications such as wide-field-of-view photonic crystals [198], although the angular dependence of the VPBGF is not explored here.

Two devices are studied in this section - one constructed from dielectric (silicon), and one constructed from a perfect electrical conductor (PEC). Since it is desired to asses the quality factor (Q) of the defect mode transmission peaks, the transmission coefficients are presented directly.

8.5.3.1 Silicon devices

The transmission intensities of the silicon device in two positional settings are presented in Fig. 8.22, with TE in Fig. 8.22(a) and TM in Fig. 8.22(b). The 0 μm shift corresponds to the middle wafers being fully interlocked, while the 130 μm shift corresponds to a separation of half the rod depth.

For both TE and TM there is a tunable defect mode associated with the lower band edge. For TE there are two peaks labelled TE_1 and TE_2, both narrow, and one has good transmission. This split peak is symptomatic of strong coupling, and being associated with the lower edge it arises from the air layers. This peak is therefore not unique to the two groups mode, and related peaks may indeed be seen in other VPBGF structures (for example Fig. 8.10 for $3.0 < n_s < 3.5$). This defect most probably arises from the top and bottom surfaces of the structure, and hence the double peak that

indicates good coupling. The TM defect mode at the lower band edge is too broad to be of use for a narrow band filter.

The more interesting feature, particularly for TE, is the defect mode transmission peak located just below 150 GHz (TE_3), and surrounded on either side by stop band. The shift in this peak is minimal, approximately 1 GHz. A similar shift is seen for the equivalent TM peak. There are some more interesting features at higher frequencies, but the simulations are not necessarily reliable, and these would need to be investigated by enlarging the structure to bring the TM peaks at around 150 GHz down to 117 GHz or below, where confidence may be had in the simulation results. Alternatively, if suitable measurement equipment was available, they could be investigated experimentally.

8.5.3.2 Metallic Devices

Metallic photonic crystals are known to reject long wavelengths [199], which would be useful for isolating the defect mode transmission resonances of the two groups structure. In this study, the metal was modelled as a perfect electrical conductor [124]. Using the same dimensions as the dielectric structure resulted in a device that only transmitted near 150 GHz. The period of the structure was doubled (giving period $\Lambda = 2.4$ mm, rod width $r = 800$ μm, but maintaining rod depth $d/2 = 262.5$ μm) to bring these transmission peaks down to a frequency range for which there was good confidence in the simulation technique. The transmission coefficients of the enlarged structure are presented in Fig. 8.23, with the TE response in Fig. 8.23(a) and the TM response in Fig. 8.23(b).

For TE, there is no peak in the fully interlocked position (0 μm), but small double-peaked transmission resonances develop when the separation is increased. Due to frequency-resolution limitations imposed by the time-step of the FDTD simulations, it is difficult to assess the peak transmission of these defects, however, looking at the inset in Fig. 8.23(a) it does not appear to approach unity.

For TM, there is a strong peak at 116 GHz at full interlock (0 μm). As the separation is increased to the rod depth (262 μm), the peak shifts down in frequency by a considerable 8 GHz. The Q of this peak, defined as the centre frequency divided by the full-width-half-maximum (FWHM), is 38 (given by 116 GHz / 3 GHz). The shift-to-FWHM ratio is 2.7 (given by 8 GHz / 3 GHz). Importantly, this peak may be shifted by an amount greater than its width. The sensitivity of the shift is 8 GHz / 262 μm, or 30 GHz/mm, although this figure is less useful since it is specific to the operating wavelength.

The metallic two-groups filter is a polarising filter with a good tuning range for the TM polarisation, and it has a narrow peak with the higher Q (38) than for the

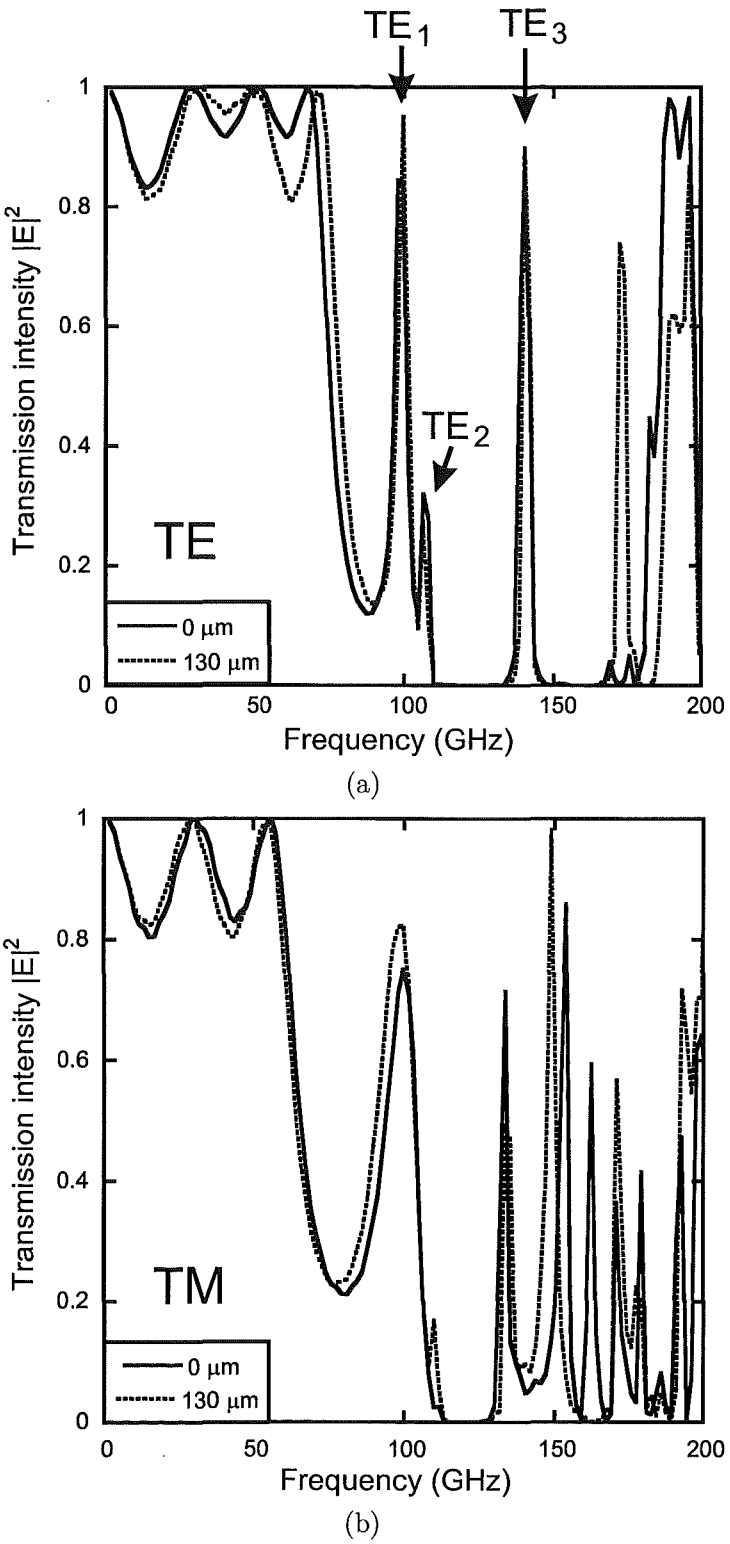


Figure 8.22 Transmission intensity of a silicon VPBGF operated in the "two groups" mode. The 0 μm shift position corresponds to full interlock of the rods in the middle layer (between the two groups), while the 130 μm shift position corresponds to a separation of one half of a rod depth. (a) TE polarisation, (b) TM polarisation.

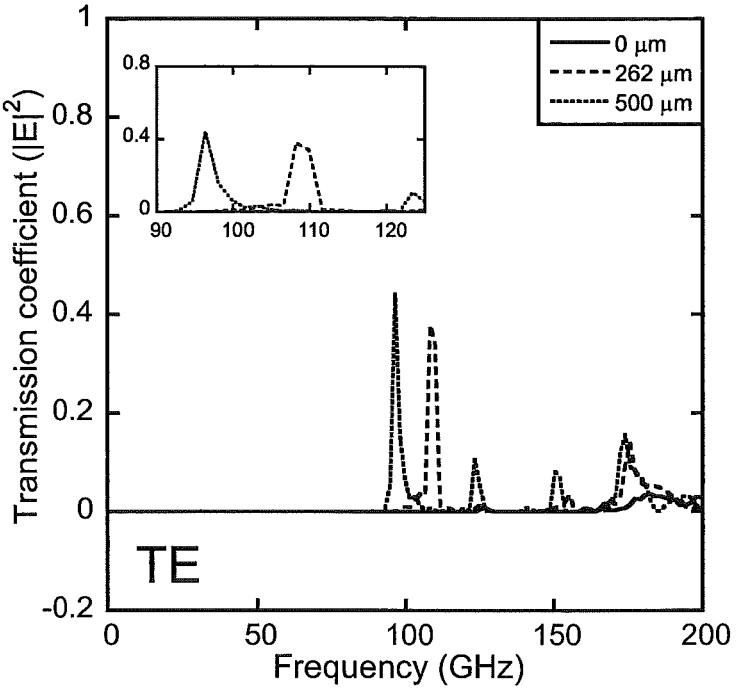


Figure 8.23 Transmission intensity of a perfect electric conductor (PEC) VPBGF operated in the "two groups" mode. The 0 μm shift position corresponds to full interlock of the rods in the middle layer (between the two groups), the 262 μm shift position corresponds to a separation of a rod depth, while the 500 μm position corresponds to a separation of two rod depths. In the 500 μm position, there is an air gap between the groups. (a) TE polarisation, (b) TM polarisation.

dielectric device. Of the devices presented so far, it would probably be the most useful as a filter. However, there is still one more mode to investigate.

8.5.4 Lateral shift (metallic structure)

The "lateral shift" mode of operation is the last to be considered here, and is shown in Fig. 8.16(e). This mode of operation has not been reported in the literature, although it is a natural conclusion of the discussions of creating disorder in photonic crystals and is therefore well founded. It bears some similarities to tunable evanescent grating couplers [200], although they are always dielectric structures and the device considered here will be constructed from metal, in one case to take advantage of a particular microcavity transmission resonance associated with metallic photonic crystals.

Since the grid is periodic, it can be represented by the unit cell of Fig. 8.24. The two plates are shown as they would be in a device in Fig. 8.24(a), while the unit cell itself is represented in Fig. 8.24(b). The actual device comprises some 400 of these unit cells, arranged in a square of 20 by 20 unit cells. For the demonstrated device, the width of the unit cell is 1.8 mm, with each rod within the unit cell being 0.4 mm wide, and 0.5 mm thick. Each plate is 1 mm thick.

The lateral shift method of tuning is further illustrated in the plan view of the unit cell shown in Fig. 8.25. The lateral spacing between the two sets of inner, interlocking rods is varied by a shift of the top plate (in this illustration, the white plate is shifted horizontally while the grey plate remains unmoved.) The position of the outer rods is not adjusted because it does not affect the tuning, although the supporting results are not presented here. It is expected that the greater the movement of the plates, the greater the shift of the narrow transmission peak. Since the shifted structure has two-fold rotationally symmetry (looking at the plan view), the maximum useful movement is a half period, $\Lambda/2 = 900$ μm. From the symmetry, the performance at any shift s_1 is the same at the equivalent shift of $s_2 = (\Lambda - s_1)$. For example, shifts of $s_1 = 200$ μm and $s_2 = 1600$ μm are equivalent in terms of the electromagnetic behaviour of this structure with $\Lambda = 1800$ μm.

The magnitudes of the simulated transmission coefficients of the two-plate metal device are shown in Fig. 8.26 for two positions. For the TE polarisation shown in Fig. 8.26(a), there is no appreciable transmission below about 170 GHz in the initial position (0 μm), and only very low levels of transmission above 110 GHz for the shifted position. Whereas, Fig. 8.26(b) shows that there are definite transmission peaks in the W-band (70 – 110 GHz). This polarising property is useful in communications systems where it is desired to double the data rate by transmitting separate channels on orthogonal (linear) polarisations. The TE polarisation would be rejected, whilst the TM polarisation would be received, and filtered. The number of wavelength division multiplexed channels that could be selected between by the device depends on the

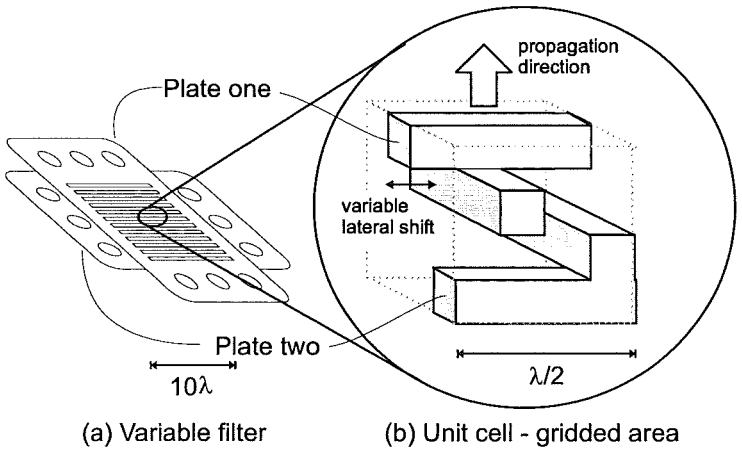


Figure 8.24 The unit cell of the gridded area of the variable filter. (a) plate orientation, (b) unit cell. The approximate size relative to the design wavelength λ, the direction of propagation, and the direction in which the plate positioning is adjusted, are indicated.

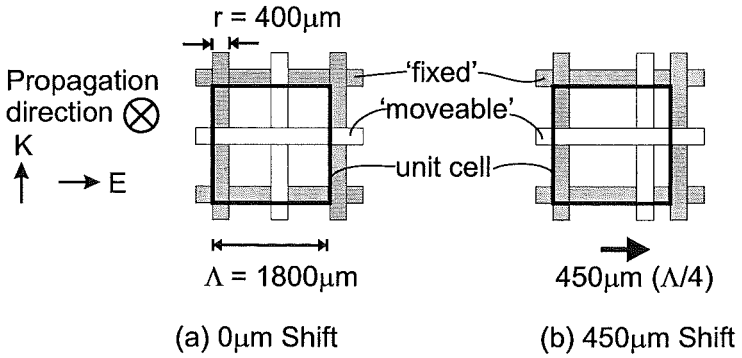


Figure 8.25 Plan view of the unit cell of the lateral shift method of tuning. (a) initial position (0 μm shift), (b) example of one of the shifted positions (450 μm shift).

nature of the TM pass bands.

In the initial position (0μm), the device exhibits a TM passband at approximately 81 GHz (TM_1). In the shifted position (450μm), the passband (TM_2) is shifted by 3.5 GHz to 77.5 GHz, giving a tuning sensitivity of 3.5 GHz/450 μm, or 7.8 GHz/mm. The 3.5 GHz tuning shift is less than the 5.6 GHz FWHM of the TM_1 peak, giving a low ratio of shift-FWHM of 0.6, approximately one quarter of the 2.7 achieved by the two-groups device. The disturbance to the regular periodicity also creates a defect-mode transmission peak at 104 GHz (TM_3). The quality factor (Q) of the peaks is revealed more clearly in the expanded-frequency-scale, log-magnitude plot of the TM transmission intensity, shown in Fig. 8.27. Both the main peak (TM_1 at 0 μm shift, and TM_2 at 450 μm shift) and the defect mode peak (TM_3) are relatively broad and therefore have low Qs. The Qs of the main peaks are 14 for TM_1 and 19 for TM_2, while the defect mode Q is slightly higher at 23. These Qs are approximately half of the equivalent two-groups device.

As for the "two-groups" device, choosing metallic construction results in the rejection of all frequencies from DC (0Hz), to the passband. The single-peak nature of the passband is due to there being only two plates (and therefore only one micro-cavity in between) in the device. If the number of plates was increased, the passband would split into several distinct peaks, with the number of peaks being equal to ($N-1$ where N is the number of plates in device [124]. There may be multi-channel communications or multi-band spectroscopy applications in which multiple tunable peaks are advantageous, however these structures are not explored here.

Whilst a "two-groups" device is predicted to give the greatest shifts and the highest Qs of the several tuning methods presented in this section, the "lateral shift" device has advantages in its simplicity of construction (only two plates required), and in the high transmission of the single tunable peak (0dB). As such, it is an ideal device with which to perform an initial experimental demonstration and validation of the VPBGF concept.

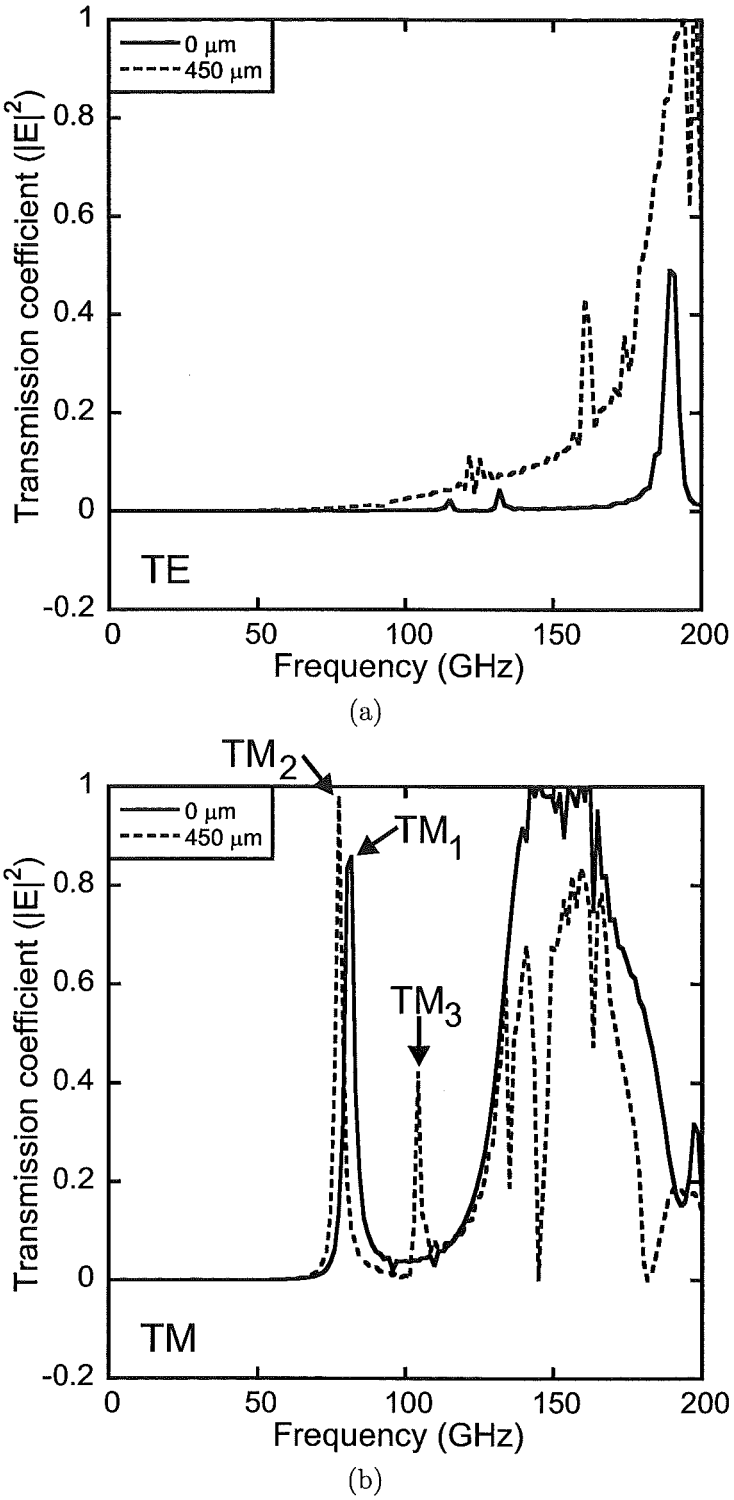


Figure 8.26 Simulated transmission intensity of a metallic VPBGF operated in the "lateral shift" mode with only two plates. The 0 μm shift position corresponds to the usual lateral alignment of the rods, while the while the 450 μm position corresponds to a lateral shift of one quarter period ($\Lambda/4$). (a) TE polarisation, (b) TM polarisation.

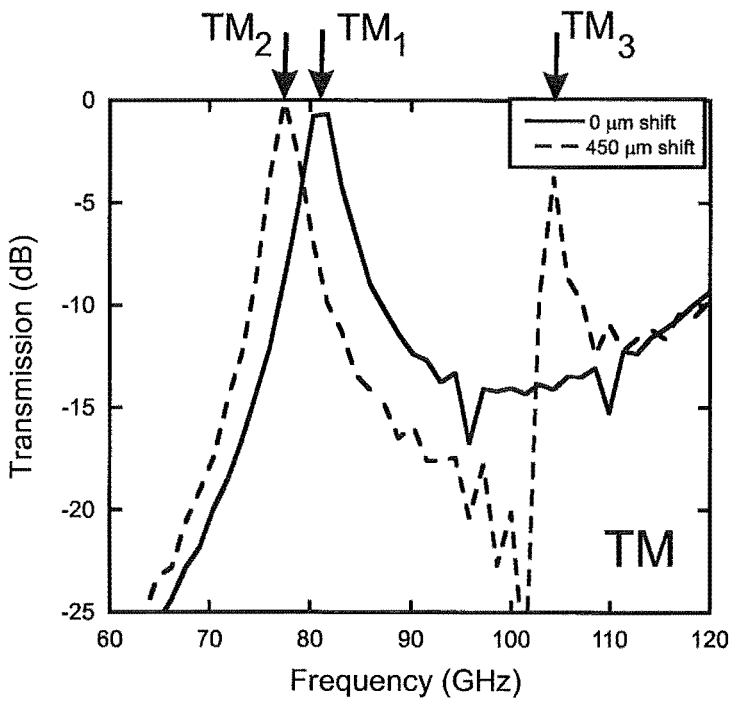


Figure 8.27 Expanded-frequency scale, log-magnitude plot of the simulated transmission for the TM polarisation of a metallic VPBGF operated in the "lateral shift" mode with only two plates. The 0 μm shift position corresponds to the usual lateral alignment of the rods, while the while the 450 μm position corresponds to a lateral shift of one quarter period ($\Lambda/4$. TM polarisation only. This is the same data as presented in Fig. 8.26(b).

8.6 LATERAL SHIFT PROTOTYPE DEMONSTRATION

A lateral shift filter was demonstrated in the W-band, using the equipment and procedures described in this section. It was constructed from two identical aluminium plates, one of which is shown in Fig. 8.28. The "operative" part of the device is the gridded area, with the surrounding solid metal surfaces present for mechanical support, mounting, and actuation purposes.

8.6.1 prototype construction

The prototype PEC lateral shift device was constructed from aluminium alloy plates (for pure aluminium, which has similar electrical properties,the room temperature conductivity is $\sigma(\mathrm{S/m}) = 3.72\times10^7$, while at 100 GHz the penetration depth is 11 μm [69]). Metals may be considered to be PEC at terahertz frequencies, because this is several orders of magnitude below the plasma frequency (in the UV). Above the plasma frequency metals become significantly lossy and dispersive, but below, they may be accurately considered as lossless PECs. One of the plates is shown in Fig. 8.28. The plates are 60 mm by 40 mm in outside dimensions, while the rods are 400 μm wide on a period of 1800 μm. These dimensions were just large enough to enable plates for a prototype to made with conventional milling. Each plate was milled from a single sheet, so that the grids were joined. Unfortunately, due to the tolerance of the milling machine ($\pm0.1mm$) and the small feature size, the rod width varied by up to 50%, far in excess of the sub-micron precision afforded by micromachining processes. Howevever, none of the grids were broken and the overall quality was sufficient for this measurement task.

8.6.2 Experimental setup

The W-band test equipment was used in a similar manner to the VADR experiments. The setup presented in Fig. 8.29 was used to measure the magnitude of the transmission coefficient of the filter at frequencies between 67 and 110GHz (W-band), in order to quantify the desired filtering action. A pyramidal horn antenna was excited by a synthesised sweeper and up-converter to produce -15dBm signal power in the W-band. A vector network analyser measured full S-parameters, from which the transmission coefficient was extracted. The transmit and receive horns were separated as widely as possible within the constraints of the equipment, by approximately 25λ, to ensure as planar wavefront as possible at the device.

A special setup was developed specifically for this experiment to allow for the correct plate alignment and separation, and enable the required lateral movements in the experiment. Each plate was fixed over a "tufnel" aperture, using the integrated mounting lugs and brass screws. One of the holders was fixed in position, while the other was actuated by a three-axis differential micrometer-driven translation stage. A

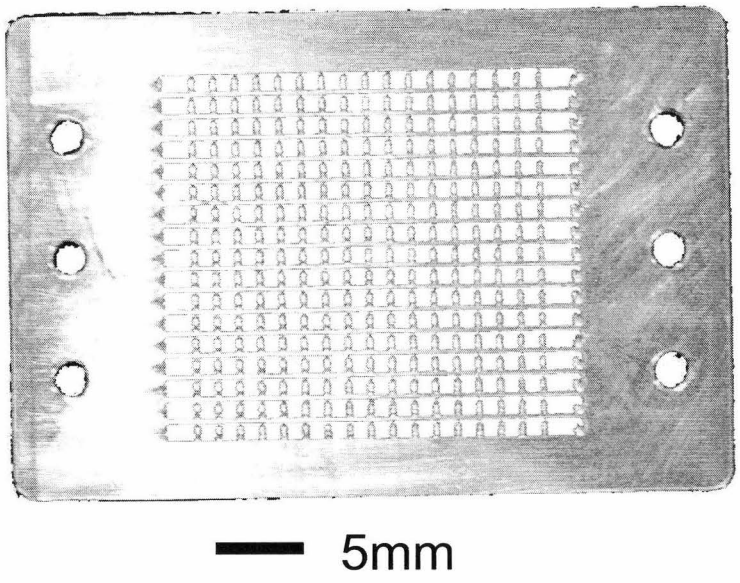


Figure 8.28 The variable filter comprises two plates like the one shown here. The plate is constructed from aluminium using conventional milling techniques, and the central area is patterned with two orthogonal grids, having a photonic band gap structure. The surrounding plate is used for mechanical support and the three holes at each end are for mounting purposes.

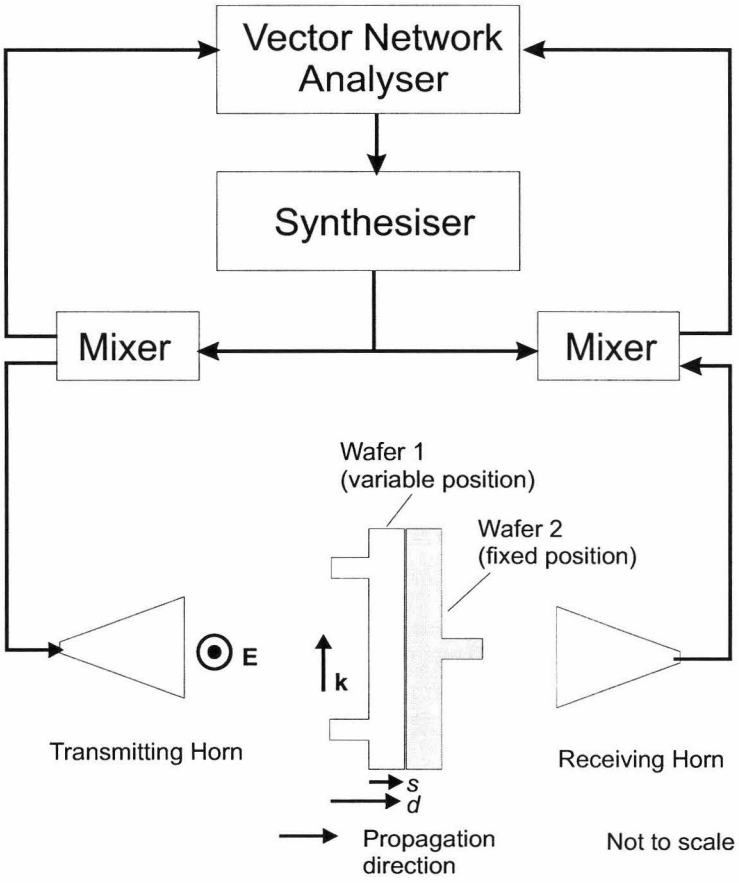


Figure 8.29 The W-band (67 - 110 GHz) measurement setup, showing the relative orientation of the horn antenna and the filter structure.

photograph of the mounting set-up is presented in Fig. 8.30. Tufnel was chosen for the holders because it was desired to minimise the number of reflective metal surfaces in the vicinity of the device, especially for surfaces that would be hard to protect with absorbing foam (such as the inside faces of the holders).

A photograph of the W-band test equipment with the VPBGF plates in position is shown in Fig. 8.31. It was necessary to raise the height of the horn antennae relative to the test platform to accommodate the VPBGF mounting arrangement. This involved placing 10 mm thick spacing blocks between under the translation stages of transmission-reflection (Tx-Rx) modules, and reversing the height-adjustable angle bracket that holds the platform to which the modules attach. A slight "droop" at the rear of the modules was corrected with 1 mm thick plastic washers. These adjustments have been made, but are not shown, in Fig. 8.31. During the experiments, extra absorbing foam is added to the top and back of the setup to enclose the device and horns and prevent stray reflections from nearby equipment. This is in addition to the foam coverings on the holders and translation stage.

8.6.3 Experimental procedure

The W-band apparatus was calibrated using the same line-reflect-line (LRL) waveguide calibration method described in Chapter 5. The entire experimental setup, including all holders and absorbing foam but excluding the VPBGF plates, was assembled and the S-parameters recorded for 256 frequencies between 67 – 110 GHz. The structure was then assembled with the plates in the TE polarisation, and S-parameters recorded at the same frequencies, for two shifts (0 and 450 μm). The plates were remounted with the TM polarisation that was expected to give the tunable peaks, and the measurements repeated.

In order to ensure that the separation distance was accurately set, on the verge of interlocking, the two plates were brought into intimate contact with slight pressure. This did not damage the grids because each plate was supported by a solid surround that would have taken much of the load. In order to change the lateral shift, the plates were drawn apart slightly to release the pressure, and therefore avoid scraping the mating surfaces or disturbing the alignment of the plates to their holders.

8.6.4 Results

The first result obtained was the transmission characteristic of the test environment with all the mounting equipment and absorbing foam in place, and this is presented in Fig. 8.32. At 70 GHz, there is 8 dB of loss in the test setup, while at 110 GHz there is 11 dB. This 3 dB difference is attributed to the greater diffraction of the beam at higher frequencies. There is 1 dB of ripple, which is attributed to unwanted reflections within the test environment.

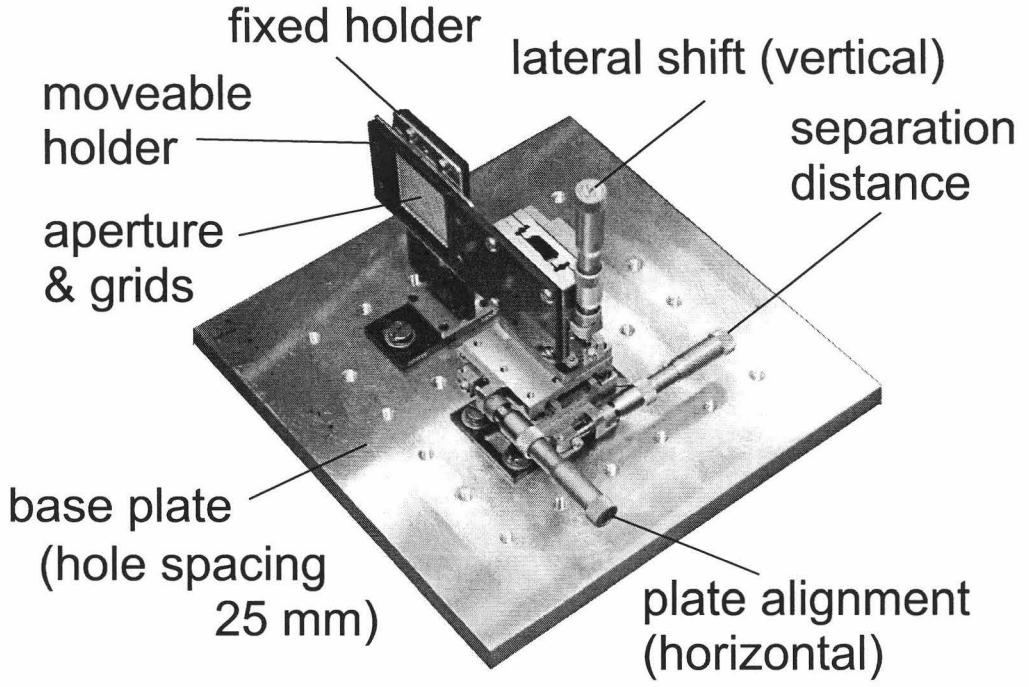


Figure 8.30 The mounting arrangement for the VPBGF lateral shift experiments in the W-band (70 – 110 GHz). Each of the two aluminium VPBGF plates is mounted onto a separate "tufnel" holder. One of the holders is fixed in position whilst the other is actuated by a three-axis differential-micrometer-driven translation stage. Each of the holders had a aperture to allow free passage of the beam.

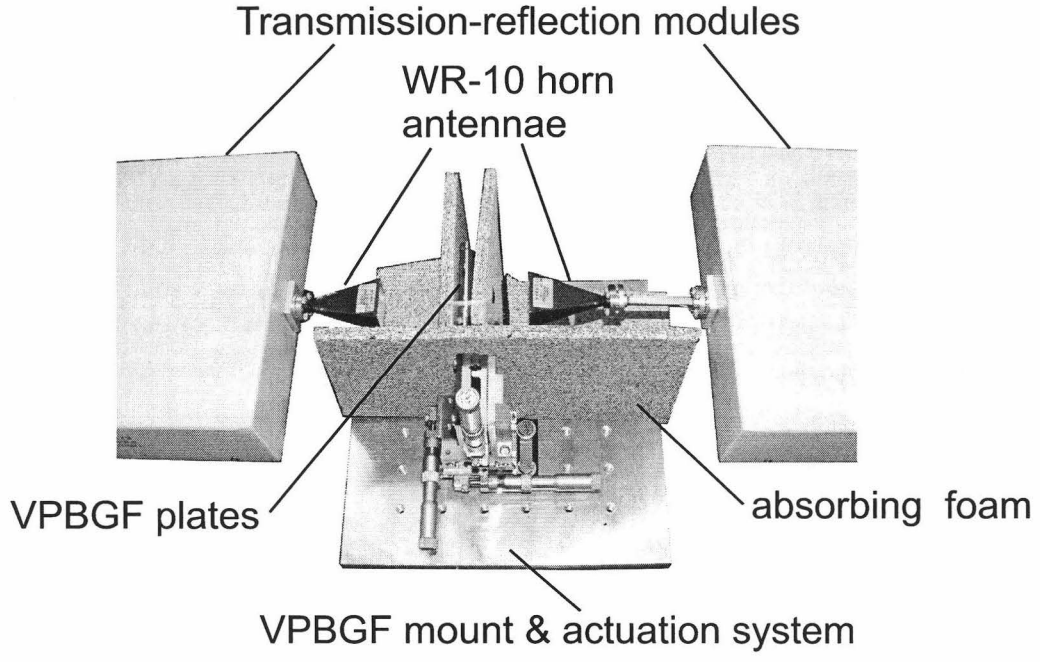


Figure 8.31 The W-band test kit with VPBGF mounted in between the horn antennae. During the experiments, extra absorbing foam is added to the top and back of the setup to enclose the device and horns and prevent reflections from the surrounding test platform (not shown).

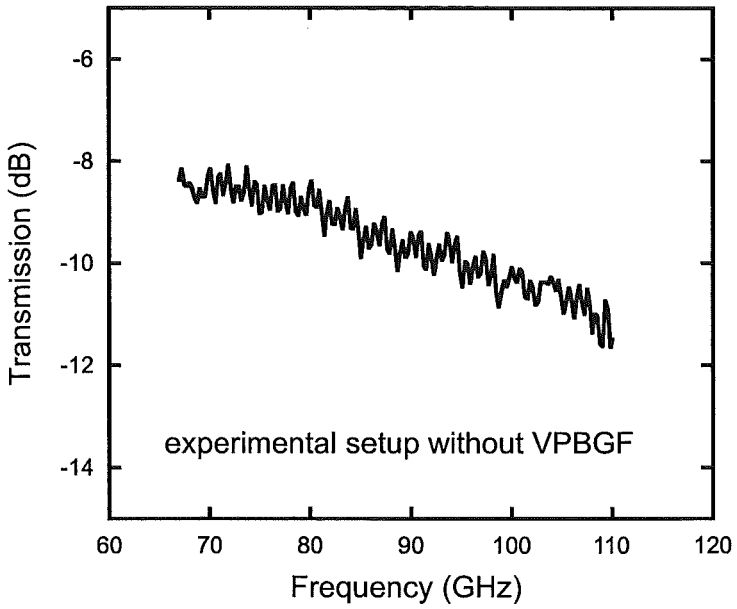


Figure 8.32 The magnitude of the measured transmission coefficient, for the test setup without the VPBGF, at frequencies of 67 – 110GHz.

Subsequently, the device was measured in the 0 and 450 μm shift positions for TM. The raw magnitude data had the test environment loss subtracted, to give the results in Fig. 8.33, with the 0 μm shift in Fig. 8.33(a) and the 450 μm shift in Fig. 8.33(b). The simulated results are plotted as well, for comparison. The measured results match the simulated behaviour, including the presence of the defect-mode resonance in the 450μm shift position. The measured peak shifts down 1.7 GHz (from 80.5 - 78.8), so the movement is not as pronounced as in the simulations (3 $\sim$ 4GHz shift). The Q of the main peak is approximately eight. A defect mode resonance is also visible in the measured results for the 450μm shift position, although it is located at 98GHz (6 GHz lower than in the simulation). The oscillations in the second stop band are most likely due to imperfections in the structure, and reflections in the test environment. The loss at the peak is 1 dB.

Plotting the two measured transmission coefficients gives a better impression of the experimentally-achieved shift; this is presented in Fig. 8.34. W-band measurements, particularly in free-space, are challenging, and the results given here represent a good degree of agreement between theory and experiment.

8.7 SUMMARY

A novel layer-by-layer photonic band gap filter (VPBGF) was proposed and investigated. The filter was tunable "on-the-fly" because it had layers of interlocking teeth, that could be moved during operation. It was shown to have potentially useful characteristics such as a tunable stop band (frequency, and width) and tunable narrow

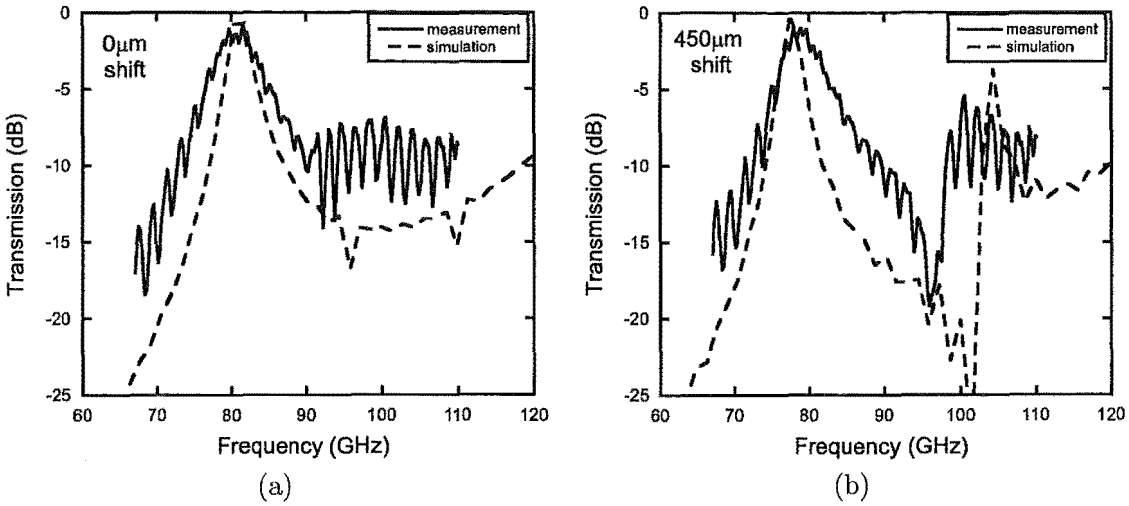


Figure 8.33 Measured TM transmission coefficient of VPBGF, plotted against the FDTD simulations. (a) 0 μm shift, (b) 450 μm shift.

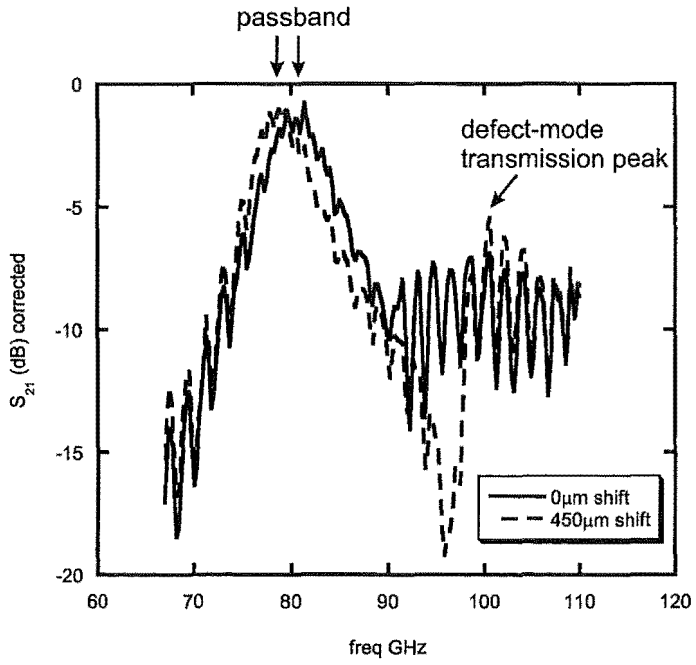


Figure 8.34 The magnitude of the measured transmission coefficient, for the variable filter at frequencies of 67 - 110GHz.

pass-bands.

The frequency of the tunable band gap and pass-bands may be located within the desired range of frequencies (such as the W-band 70 – 110 GHz) by adjusting the fabrication parameters of the plates that comprise the device. Increasing the effective structure size, relative to the material wavelength, caused the band gap to drop in frequency as expected.

Several schemes for setting and varying the plate separation were presented, and each provided a unique method of tuning. The most successful were the two groups method, and the lateral shift method, both of which produced narrow tunable pass band peaks. The two groups method had a higher Q (38) than the lateral shift mode (19). Since the latter required less plates, it was investigated experimentally in the W-band, and the measurements were shown to have good agreement with the simulations. Since the same simulation technique was used throughout the chapter, this gives confidence in the other predictions.

Further work would be required before this device could be shown to represent a definite improvement over existing filter technologies in all areas, particularly in terms of the shift to full-width-half-maximum ratio (best to date is predicted to be 2.7), however the simple fabrication, compact structures and potential for high rejection are favourable points of this design.

Chapter 9

CONCLUSIONS AND FURTHER WORK

The study presented in this thesis investigated two types of passive device for terahertz frequencies (100 GHz – 10 THz). First, an existing concept for a variable artificial dielectric retarder (VADR) was experimentally demonstrated and improved devices proposed. Second, a new tunable woodpile-like photonic crystal was proposed (VP-BGF) and experimentally demonstrated. The main achievements of this work may be listed as follows:

- developed routinely successful microfabrication procedure for patterning the VADR device's 20 by 20 mm silicon plates with 500 μm-period V-grooves, including photolithographic mask alignment without wafer flats to within $\pm 1^\circ$ of the $< 110 >$ directions;

- experimentally demonstrated the VADR device at the design frequency of 100 GHz, and surrounding W-band frequencies (70 – 110 GHz), representing the first demonstration of a variable polarisation compensator using artificial dielectrics, and the first demonstration of a variable polarisation compensator designed specially to operate at terahertz frequencies;

- showed with rigorous full-vector electromagnetic simulations (rigorous coupled-wave analysis) that the 100 GHz silicon VADR prototype performed according to the principles described in the original proposal, despite deliberate omission of the anti-reflection coatings (the as-yet uncharacterised anti-reflection coatings were omitted from the prototype to reduce uncertainty in the measurements);

- proposed two new VADR devices with greater retardance and reduced insertion loss (R-VADR: quarter wave, 2.6 dB, respectively; R^2-VADR: half wave, 0.6 dB);

- proposed a third new VADR device (DARTH-VADR) that overcomes the practical difficulty of achieving zero retardance by not requiring full interlock of the plates, that is also low loss (0.6 dB), capable of quarter-wave retardance, and requires only a single movable plate;

- proposed a new type of tunable woodpile-like photonic crystal (named "variable photonic band gap filter," or VPBGF) with easily accessible external control surfaces, and a huge number of possible tuning modes;
- experimentally demonstrated in the W-band (70 – 110 GHz) a metallic VPBGF in good agreement with rigorous full-vector electromagnetic simulations (finite-difference time-domain);
- all VADR and VPBGF devices scale to operate at any terahertz frequency whilst maintaining the same high levels of performance, subject to fabrication and alignment limitations.

The following sections summarise the work supporting these achievements, and place them in the wider context of the emerging terahertz technologies.

9.1 VADR

The VADR device was proposed by Cumming and Blaikie [65], and is the first variable polarisation compensator to use artificial dielectrics and the first designed to operate at terahertz frequencies. As explained in chapter two, many existing polarisation compensation techniques for visible frequencies employ naturally birefringent materials, such as calcite, which exhibit poor properties at terahertz frequencies, whereas the VADR device is an attractive and compact alternative.

A 100 GHz silicon prototype VADR device was a scaled version of the originally-proposed 1 THz VADR device, highlighting the ease with which the VADR design may be scaled to new operating frequencies. It was chosen to fabricate the device in silicon because of its low loss and low dispersion at terahertz frequencies (true for high resistivity silicon). The design frequency was chosen to coincide with the W-band range (70 – 110 GHz) of an available state-of-the-art vector network analyser and associated test equipment. Measurements at the lower end of the terahertz frequency range were still challenging, but promised to provide the detailed and accurate results required to further develop the VADR device.

Unfortunately, this made the microfabrication task more difficult due to the large feature size required. While conventional machining and milling techniques were not suitable because of the high precision required, the V-grooves were ideal for fabrication with anisotropic alkaline (KOH) etching of silicon. Difficulties arose because accurate rotational alignment of the photolithographic mask to silicon crystal planes was critical to ensure the grooves were not destroyed by excessive mask undercut during the long wet etch ($\sim$ 260 min). With the available microfabrication equipment it was not possible to perform photolithography or etching on a whole 100 mm silicon wafer at once. As a consequence, each VADR plate had to be microfabricated separately, without

the aid of the wafer flats for mask alignment. Instead, a short pre-etch with a special "dual radial alignment mark" was performed individually on each sample, to which the final mask for the V-grooves could be aligned to within $\pm 1^\circ$ of the $< 110 >$ crystal directions.

The final microfabrication procedure gave routine success with nine plates being produced with no failures. Nine plates was enough for four devices and a spare, and the best two plates were selected for measurement. The original proposal [65] suggested using anti-reflection (AR) coatings on the outside (unpatterned) surfaces of the VADR device to suppress reflections. No AR coatings were applied to the prototype, in order to reduce the number of aspects of the device under test and avoid additional uncertainties in the measured results.

Since the device is quasi-optical, a free-space measurement system was required. Just such a system was available at the University of Glasgow, comprising vector network analyser, sweep synthesiser and mixer units on which horn antennas could be mounted. Special mounting equipment was devised for the VADR plates in order to ensure good alignment and accurate control of plate separation, using a combination of off-the-shelf optical components, such as a translation stage and custom components, such as the fibreglass apertures over which of the plates was mounted. In order to completely characterise the device, the key figures of merit (retardance, insertion loss) were obtained by measuring the TE and TM complex transmission coefficients. The measured maximum retardance was less than a quarter wave ($\Gamma_{max} = 74^\circ$), with a sensitivity of 120°/mm in the first 100 μm of the interlock regions. Insertions losses were as high as 7 dB for TE and 11 dB for TM. While these results were in approximate agreement with the predictions of the original proposal [65], some discrepancies were evident.

These discrepancies were not unexpected, because the AR coatings had been deliberately omitted on the prototype. In order to show that the discrepancies were solely due to the lack of AR coatings, it was necessary to perform further simulations. The assumption of AR-coatings had simplified the proposal's simulation model [65] because it was not necessary to account for any reflections off the outside surfaces of the device. In order to model a VADR device with no AR-coatings, further boundary conditions were required but their addition was inconvenient in the originally-employed multiple multipole (MMP) Maxwell equation solver tool. Instead, an alternative Maxwell equation solver tool (Gsolver) was used, based on rigorous coupled wave analysis (RCWA). The complex refractive index of the p-type $10 - 20$-Ω-cm silicon used in the VADR prototype was deduced by fitting the simulated performance of the etalon to the measured performance, giving $n_s = 3.41 + 0.257i$. Confidence in this value was good because it lay between literature values measured for p-type 9.0 Ω-cm and 50 Ω-cm silicon. Unfortunately, a value for 20 Ω-cm silicon (the upper bound of the wafer's specified resistivity) could not be found in the literature. The Gsolver model of VADR approxi-

mated the grooves with a ten layer stepped-staircase. Simulated results for the TE and TM transmission coefficients were in good agreement with the measurements. Specifically, the transmission coefficients were predicted to within 2 dB and 11° in the worst case, and the maximum retardance was predicted to be 82°, within 8° of the value calculated from the measurements (74°).

With good confidence in the RCWA simulation technique, it was possible to investigate improvements to the VADR device. Three aspects required attention. Firstly, the high insertion loss could be a problem with the reasonably low signal strengths expected from terahertz sources. In order to reduce the loss, preferably to 1 dB or less, higher resistivity silicon (5 kΩ-cm) may be used to eliminate absorption in the substrate [181], while AR coatings may be used to reduce reflections from the outer surfaces of the device [65]. Secondly, the maximum retardance is less than a quarter wave and reduces further when AR coatings are added to reduce the loss. The ability to produce a quarter wave retardance enables a number of useful polarisation transformations. Thirdly, it is difficult to achieve zero retardance in practice because of mechanical imperfections in the V-grooves that prevent the plates from fully interlocking. The ability to produce zero retardance is of particular importance when the device is used to provide minor corrections over time - sometimes no correction is needed, and, in non-laboratory applications, it would be impractical to remove the device at such times. To solve these problems, three new VADR devices were proposed, all of which would use the lossless high-resistivity substrate and the AR coatings. The three new devices are shown in Fig. 9.1, comprising the original VADR for comparison in Fig. 9.1(a), the single rectangular grating-pair device (R-VADR) in Fig. 9.1(b), the double rectangular grating-pair device (R^2-VADR) in Fig. 9.1(c) and the dual-axis rectangular-tooth half-wave device (DARTH-VADR) in Fig. 9.1(d).

Rectangular grooves were used in the new devices because the increased birefringence gives greater retardance for the same groove depth g. The R-VADR is capable of over a quarter wave retardance ($\Gamma_{max} = 112°$) for the same groove depth g as the VADR, with approximately half the insertion loss (2.6 dB). The R^2-VADR device is essentially two cascaded R-VADRs, with the two inner plates combined. Since the two grating pairs are parallel, the retardances realised in each add up to give a greater total retardance than a single grating device. An example device had a maximum retardance well in excess of a half wave ($\Gamma_{max} = 416°$), and an insertion loss of 0.6 dB. The R^2-VADR offers the full range of retardances, with an attractive insertion loss less than 1 dB, and would make an excellent device for applications requiring quarter and half wave transformations.

The third device, the DARTH-VADR, is again essentially two cascaded R-VADRs, but the grating pairs are oriented orthogonally so that they work in opposition. It is actuated by holding the outer plates still and moving the centre plate. When the centre plate is centred, the separation for each grating pair is the same ($s_1 = s_2$) and

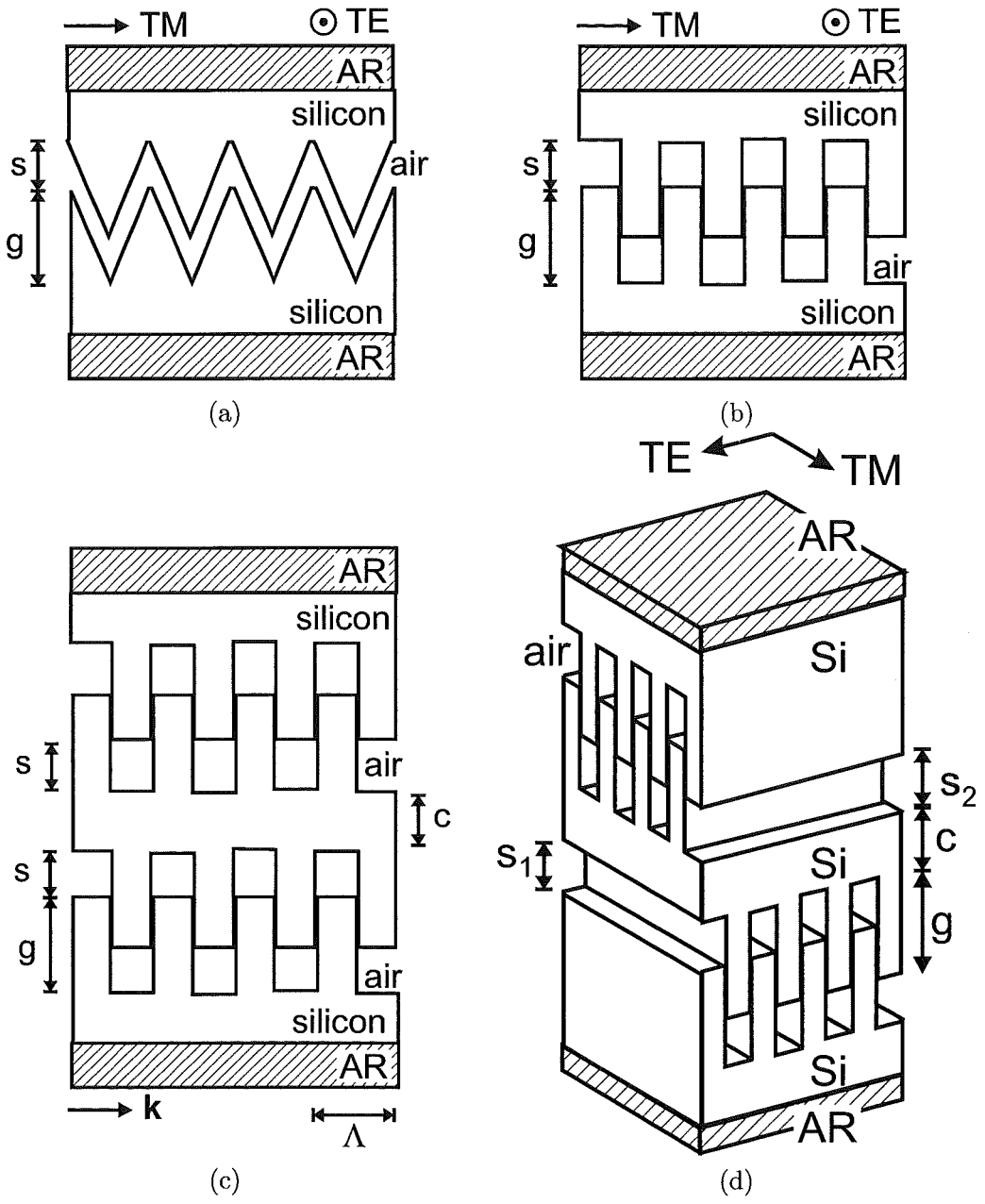


Figure 9.1 The original, and the three new, VADR devices. The dimensions are groove depth d, separation s and centre-thickness c. (a) VADR (original) (b) R-VADR (new), (c) R^2-VADR (new), (d) DARTH-VADR (new).

the retardances generated by the individual grating pairs cancel. Thus, zero retardance is achieved without requiring full interlock of the plates. The actuation scheme would be ideal for applications where small, fast, corrections are required to correct minor deviations from a particular state of polarisation. It would be equally useful for quarter wave transformations. The device has half-wave in the name because it is capable of both a negative and positive quarter wave shift (depending on which direction the centre plate is displaced). With a low insertion loss of 0.6 dB, the DARTH-VADR meets the goals for improvement of the VADR device, and draws this aspect of the study to a close.

9.1.1 Future work

Compared to the performance of traditional compensators at terahertz frequencies, the R^2-VADR and DARTH-VADR are particularly attractive devices, and their development has reached the stage where they are ready to optimised for a specific application. Since terahertz technology is still at an early stage, there is no immediate application for these devices that the author is aware of. When such time comes, the following aspects of the VADR devices will require attention. The microfabrication process could be improved by processing entire wafers at once and including integral mounting lugs around the edges of the gratings. The method of actuation in an application will to a certain extent depend on the speed at which it is desired to make adjustments. The present system of manually adjusting a micrometer is sufficient for measurement purposes, and would be fine in a laboratory situation. Piezo-electric actuators appear promising for commercial or industrial applications [201], such as using the DARTH-VADR to dynamically compensate for deviations from a desired polarisation state. Packaging is another issue, as is operation in harsh environments. For example, exposure to an alkaline atmosphere would soon degrade the silicon. However, the device would work equally well with other materials, the only proviso being that a high refractive index material would allow a more compact device to be created. Even with the precision of silicon microfabrication, fabrication of devices for use beyond the upper bounds of the terahertz part of the spectrum would be difficult, due to the precise alignment required between layers. The complete and thorough investigation of the VADR devices contained in this thesis provides an excellent starting point for tackling these "real world" issues.

9.2 VPBGF

Inspired by the DARTH-VADR design, but unrelated in purpose, the variable photonic band gap filter (VPBGF) investigated in the penultimate chapter of this thesis is a new type of tunable photonic crystal. As shown in Fig. 9.2 The VPBGF bears some

similarity in appearance to the layer-by-layer or "woodpile" photonic crystals, but these existing photonic crystals do not have interlocking teeth, or the easily accessible external control surfaces that allow individual plates to be actuated at will. These allow a huge number of tuning modes to be realised and gives great control over the number and type of defects.

In order to investigate these structures theoretically, a finite-difference time-domain Maxwell equation solver tool was adopted. This tool was better equipped than the previously-used Gsolver RCWA to handle the crossed-gratings and larger feature sizes (the VPBGF grating period is at least twice that of the VADR devices). The effects of the plate fabrication parameters such as grating period, filling factor and refractive index contrast were shown by simulation to exhibit the trends already reported in the literature for such structures. Essentially, increasing any dimension with respect to the wavelength causes the band gap to drop in frequency. This is important when one or more parameters are constrained. For example, the rod depth might be fixed by the wafer thickness, but, the grating period may be adjusted at the mask design stage in order to force the band gap to appear at the desired frequency. For example, a six-plate silicon device fabricated from 525 μm thick silicon wafers can be forced to exhibit a band gap in the F-band (120 – 150 GHz) by setting the period to 1.2 – 1.4 mm, or in the W-band (70 – 110 GHz) by setting the period to 1.5 – 1.8 mm.

An initial investigation into the huge number of possible tuning methods covered most of the more apparent methods, for both silicon and metal devices. Two methods appeared superior, the "two groups" and the "lateral shift" mode. With these modes, a tunable narrow peak is produced. A metal device tuned with the "two groups" method is predicted to produce the best quality factor (Q), of 38. The ratio of the peak's shift to its full-width half-maximum (FWHM) is 2.7, which means that the peak shifts further than its width. However, this peak is not yet narrow enough to switch between densely packed wavelength-division-multiplexed channels, but could be of use in switching between bands in an imaging or spectroscopy application. Alternatively, it could be used in conjunction with a broad band source or detector to provide a compact, simple, tunable device.

The second best method was the "lateral shift", with approximately half the Q (14 – 19). This device takes advantage of a passband just above the low frequency cutoff in a metallic photonic crystal, for the TM polarisation. When there are only two plates, the TM passband has a single peak with 100% transmission. By relative lateral shifts of the two plates, the position of this pass band may be tuned. This device is completely different to metal mesh filters for the infrared, because they form an etalon that is tuned by adjusting the separation between the plates. Due to the simple construction of the "lateral shift" VPBGF device, it was selected for an experimental demonstration in the W-band. The two plates were constructed from aluminium using conventional machining and milling techniques. The plates were 60 mm by 40 mm, the

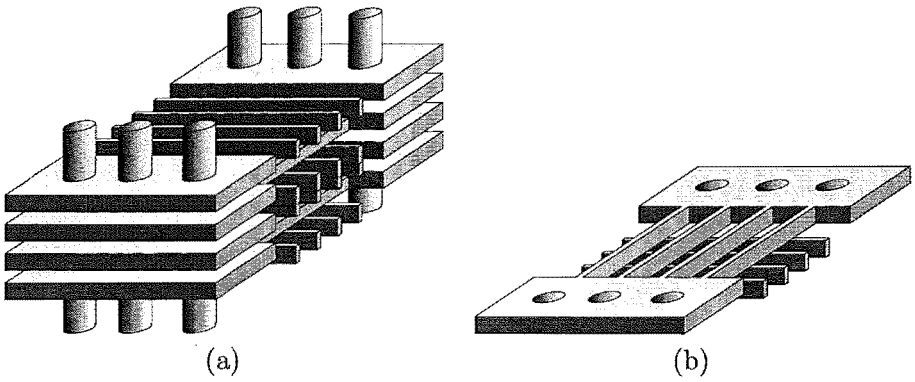


Figure 9.2 The VPBGF device comprises a stack of plates (a), where each plate comprises two orthogonal gratings (b). One of several possible methods of actuating the layers involves threading the plates onto a series of rods, some of which will move the plates, others which will simply keep the stack aligned.

rods were 400 μm wide on a period of 1.8 mm and 0.5 mm thick. The TM transmission coefficient was measured for frequencies 70 – 110 GHz for two positions, and found to be in good agreement with the simulations. The measured shift of the peak was 1.7 GHz (80.5 to 78.8 GHz) for a lateral plate movement of 450 μm. The measured Q of 8 was approximately half that predicted and quite modest for a narrow passband. While further work would be required before this filter device would represent a definite improvement over existing filter technologies in all areas, the simple fabrication, compact structures and potential for strong stopband rejection are favourable points of the VPBGF design.

9.2.1 Future work

Suggestions for further work on the VPBGF device include investigating improvements to the measurement technique, as well as measurements of further prototypes. The presence of reflections between the horn antennas and the device manifests itself as a ripple in the measured results. The quality of the results could be improved by taking a number of measurements with the horn antennas at different spacings and averaging the measurements. While the lateral shift device results were an encouraging validation of the simulation methods, it would be worth investigating the "two groups" metal device as that is predicted to have the highest Q of all the VPBGF devices investigated so far, as well as the greatest shift to FWHM ratio. Aspects of investigations into wide-field-of-view filters suggest that the VPBGF structure might be able to be modified to become a polarisation insensitive version [202]. If so, the structure would be of use in wide-field-of-view imaging applications at millimetre-wave and terahertz frequencies.

The worldwide effort at developing sources, devices and systems for terahertz frequencies involves considerable effort by numerous researchers. While great progress

has already been made (commercial terahertz imaging systems are available from Picometrix [203] and TeraView [204]), there is still a long way to go before it could be said that terahertz frequencies are fully exploited. However, if the present rapid rate of new developments relating to terahertz frequencies continues, then it bodes well for the future.

Appendix A

ADDITIONAL BACKGROUND MATERIAL

A.1 THE ONE-DIMENSIONAL WAVE EQUATION

The one-dimensional wave equation may be derived from Maxwell's equations using the following method [69]. It may then be solved to give the equation of a uniform plane wave. The resulting equation is referred to in section 2.2.1.

In a homogeneous region with constant, scalar permittivity ε and permeability μ, and with no free charges and currents, Maxwell's equations are:

$$\nabla \cdot \mathbf{B} = 0 \tag{A.1a}$$

$$\nabla \cdot \mathbf{D} = \rho \tag{A.1b}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} = -\mu \frac{\partial \mathbf{H}}{\partial t} \tag{A.1c}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J} + \varepsilon \frac{\partial \mathbf{E}}{\partial t} \tag{A.1d}$$

where $\mathbf{D}$ is the electric flux density, $\mathbf{B}$ is the magnetic flux density, $\mathbf{E}$ is the electric field, $\mathbf{H}$ is the magnetic field, ρ is the total charge, and $\mathbf{J}$ is the conduction (or convection) current, with the use of the bold face denoting a vector quantity.

In a three dimensional Cartesian co-ordinate system, where the x, y, z axes lie along the directions of the unit vectors $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$, the electric and magnetic fields can be broken down into three components:

$$\mathbf{E} = E_x \hat{\mathbf{x}} + E_y \hat{\mathbf{y}} + E_z \hat{\mathbf{z}} \tag{A.2a}$$

$$\mathbf{H} = H_x\hat{\mathbf{x}} + H_y\hat{\mathbf{y}} + H_z\hat{\mathbf{z}}. \tag{A.2b}$$

The one dimensional wave equation may be developed by setting the direction of propagation to be z, and, since the wave is uniform, assuming that the only variation in the fields occurs in the direction of propagation. Thus, the partial derivatives in the non-varying directions will be zero, $\partial/\partial x = 0$ and $\partial/\partial y = 0$. With these conditions, the two curl equations, Equations A.1c and A.1d, can be broken into component form as follows:

$$\frac{\partial E_y}{\partial z} = \mu\frac{\partial H_x}{\partial t} \tag{A.3a}$$

$$\frac{\partial E_x}{\partial z} = -\mu\frac{\partial H_y}{\partial t} \tag{A.3b}$$

$$0 = \mu\frac{\partial H_z}{\partial t} \tag{A.3c}$$

$$\frac{\partial H_y}{\partial z} = -\varepsilon\frac{\partial E_x}{\partial t} \tag{A.3d}$$

$$\frac{\partial H_x}{\partial z} = \varepsilon\frac{\partial E_y}{\partial t} \tag{A.3e}$$

$$0 = \varepsilon\frac{\partial E_z}{\partial t} \tag{A.3f}$$

Equations A.3b and A.3d may be combined if, first, the partial derivative of Equation A.3b is taken with respect to z

$$\frac{\partial^2 E_x}{\partial z^2} = -\mu\frac{\partial^2 H_y}{\partial z \partial t}, \tag{A.4}$$

and the partial derivative of Equation A.3d is taken with respect to t

$$\frac{\partial^2 H_y}{\partial z \partial t} = -\varepsilon\frac{\partial^2 E_x}{\partial t^2}. \tag{A.5}$$

Then, Equation A.5 may be substituted into Equation A.4, giving the desired one-

dimensional wave equation in E_x

$$\frac{\partial^2 E_x}{\partial z^2} = \mu\varepsilon \frac{\partial^2 E_x}{\partial t^2}. \tag{A.6}$$

According to Reference [69], the general solution to the wave equation is

$$E_x(z,t) = g_1\left(t - \frac{z}{v}\right) + g_2\left(t + \frac{z}{v}\right), \tag{A.7}$$

where t is time, g_1 and g_2 are arbitrary functions and

$$v = \frac{1}{\sqrt{\mu\varepsilon}} \tag{A.8}$$

is the velocity with which the waves propagate in the z direction. In free space,

$$v = c = \frac{1}{\sqrt{\mu_0\varepsilon_0}} \tag{A.9}$$

where $\mu_0 = 4\pi \times 10^{-7}$, $\varepsilon = 8.854 \times 10^{-12}$ and $c = 2.998 \times 10^8$. A special solution of the wave equation is the sinusoidal wave, of the form

$$E_x(z,t) = A_1 \sin\omega\left(t - \frac{z}{v}\right) + A_2 \sin\omega\left(t + \frac{z}{v}\right) \tag{A.10}$$

where A_1 and A_2 are the electric field amplitudes of the waves travelling in the positive and negative z directions respectively , and $\omega = 2\pi f$ is the angular frequency of the wave, where f is the frequency in Hertz. The wavenumber k, is a scalar constant that represents the phase change in the wave per unit length, or metre:

$$k = \frac{\omega}{v} = \omega\sqrt{\mu\varepsilon}. \tag{A.11}$$

The distance in the z direction in which the phase changes by 2π is the wavelength λ,

$$\lambda = \frac{2\pi}{k}. \tag{A.12}$$

The wavelength is related to the frequency, and in free space is given by

$$\lambda = \frac{v}{f}. \tag{A.13}$$

In other materials, the wavelength differs due to the different permittivity, permeability, or both. For the dielectric materials considered in this thesis, the permeability is assumed to remain constant at the free space value unless otherwise specified. This assumption is made because it is unusual to find dielectrics at high frequencies with permeability $\mu \neq \mu_0$ [87]. Therefore, only the permittivity need be explicitly accounted for. In this regard, the refractive index n is a useful parameter, and is defined for a

medium as

$$n = \frac{c}{v} = \sqrt{\frac{\mu}{\mu_0}\frac{\varepsilon}{\epsilon}}, \tag{A.14}$$

and for materials such as dielectrics that have unity permeability, the expression simplifies to

$$n = \sqrt{\varepsilon_r}. \tag{A.15}$$

The wavelength in a material with refractive index n is thus:

$$\lambda_n = \frac{\lambda_0}{n}. \tag{A.16}$$

For lossy materials, the refractive index is complex:

$$\tilde{n} = n - i\kappa; \tag{A.17}$$

where $i = \sqrt{-1}$ and κ represents the extinction coefficient of the material. Unless it is stated otherwise, n will be written for convenience and can be taken to mean either n or $\tilde{n}$. Note that for complex refractive indices

$$n^2 - \kappa^2 = \varepsilon' \tag{A.18a}$$

$$2n\kappa = \varepsilon'' \tag{A.18b}$$

where $\varepsilon = \varepsilon' - i\varepsilon''$. Note that in the case of a lossless material ($\kappa = 0$), Equation A.18a equates to Equation A.15.

Important implications of Equations A.3 and A.10 are that **E** and **H** are perpendicular to each other and the direction of propagation, are in phase, and are related in magnitude by

$$\frac{E}{H} = \eta = \sqrt{\frac{\mu}{\epsilon}}, \tag{A.19}$$

where η is defined as the intrinsic impedance of the medium. The drawing in Figure A.1 serves to illustrate these relationships between the electric and magnetic fields in a simple uniform plane wave. This is important because it illustrates the fact that only one field needs to be known in order to work out the other. By convention, usually only the electric field is written.

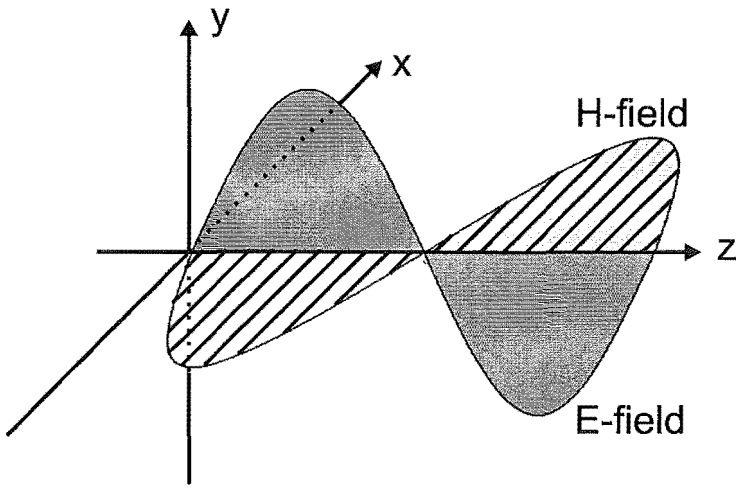


Figure A.1 A uniform electromagnetic plane-wave travelling the positive z direction, with the electric and magnetic fields plotted along the z axis, at an instant in time. The relative magnitudes of the electric and magnetic fields are related by the impedance of the medium (see Equation A.19), and are not shown to scale here.

A.2 STOKES PARAMETERS AND THE JONES CALCULUS

Several systems exist for the description of polarisation states. For example, Stokes parameters are useful for describing partially polarised waves, where one part of the wave is completely polarised (such as that emitted from an antenna) and the other part is completely unpolarised (such as that emitted from a celestial source); whereas the Jones calculus provides certain advantages in calculations of the polarisation states in systems where all radiation is monochromatic, and completely polarised. Both the Stokes parameters and the Jones calculus are introduced here, beginning with Stokes parameters.

Following Kraus [205], the Stokes parameters I, Q, U and V are calculated from the angles in the Poincaré sphere: the phase delay ϕ, the tilt angle θ and the ellipticity ξ (see Fig. 2.10), as follows:

$$I = S = S_x + S_y = \frac{< E_1^2 >}{Z} + \frac{< E_2^2 >}{Z}, \tag{A.20a}$$

$$Q = S = S_x - S_y = \frac{< E_1^2 >}{Z} - \frac{< E_2^2 >}{Z}, \tag{A.20b}$$

$$U = \frac{2}{Z} < E_1 E_2 \cos\phi >= S < \cos 2\xi \sin 2\theta >, \tag{A.20c}$$

$$V = \frac{2}{Z} < E_1 E_2 \sin\phi >= S < \sin 2\xi >, \tag{A.20d}$$

where S is the total Poynting vector magnitude for the wave ($\mathbf{S}_{av} = \frac{1}{2}\hat{\mathbf{z}}\frac{E_1^2+E_2^2}{Z_0}$), S_x is the Poynting vector component of the wave polarised in the x direction, S_y is the Poynting vector component of the wave polarised in the y direction, E_1 is the amplitude of the electric field component of the wave polarised in the x direction, E_2 is the amplitude of the electric field component of the wave polarised in the y direction, and Z is the intrinsic impedance of the medium. The angle brackets indicate the time average (for example, $< E_1^2 >= \frac{1}{T}\int_0^T [E_1(t)]^2 dt$.)

For a completely unpolarised wave, there is equal power in Poynting vector component ($S_x = S_y$), and E_1 and E_2 are uncorrelated. Thus, the Stokes paramters are

$$\begin{aligned} I &= S \\ Q &= 0 \\ U &= 0 \\ V &= 0 \end{aligned} \tag{A.21}$$

where S is the Poynting vector of the wave. It is a requirement of a completely unpolarised wave that $Q = U = V = 0$.

It may be shown by considering a number of polarisations states that I is always equal to the total power of the wave, Q is always equal to the power in the linearly polarised components of the wave in the x or Y directions, U is always equal to the power in the linearly polarised components at tilt angles of $\theta = 45°$ or $135°$, and V is equal to the power in the circularly polarised components (left- or right-handed). The Stokes parameters are often normalised by dividing each parameter by S, the total power in the wave, giving $s_0 = I/S = 1$, $s_1 = Q/S$, $s_2 = U/S$, $s_3 = V/S$. The degree of polarisation d is the ratio of completely polarised power to the total power, given by

$$d = \frac{\sqrt{Q^2 + U^2 + V^2}}{I} = \sqrt{s_1^2 + s_2^2 + S_3^2}, 0 \leq d \leq 1. \tag{A.22}$$

When the wave is completely polarised, $d = 1$, then the Jone's calculus is of use. The Jones calculus is a system of matrix notation that aids in the calculation of the output waveform after transmission through one or more optical elements such as polarisers or retarders [71].

The system may be illustrated in the following way. Consider a beam of quasi-monochromatic light travelling in the positive z direction of a right-handed Cartesian co-ordinate system. The electric field component along the x-axis may be expressed as

$$E_x = E_{x_0} \cos\left(\omega t - \frac{2\pi z}{\lambda} + \partial_x\right) \tag{A.23}$$

which is the real part of

$$E_{x_0} e^{j(\omega t - (2\pi z/\lambda) + \partial_x)} \tag{A.24}$$

and along the y-axis as

$$E_y = E_{y_0} \cos\left(\omega t - \frac{2\pi z}{\lambda} + \partial_y\right) \tag{A.25}$$

which is the real part of

$$E_{y_0} e^{j(\omega t - (2\pi z/\lambda) + \partial_y)}. \tag{A.26}$$

This notation may be made more compact if exponentials are used to describe the waves, and matrix notation is used to describe the interactions of the optical elements. The particular matrix notation is known as the "Jones calculus". The elementary Jones vectors that are used to describe the action of optical elements are given in Table A.1

The devices described in the table are as follows. The perfect polariser only trans-

Table A.1 Elementary Jones Vectors [71]

Operation	Matrix	Notes
Perfect polariser	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	
Partial polariser	$\begin{bmatrix} \tau_x & 0 \\ 0 & \tau_y \end{bmatrix}$	τ_x, τ_y are transmission coefficients
Pure Retarder	$\begin{bmatrix} 1 & 0 \\ 0 & e^{j\Gamma} \end{bmatrix}$	Γ is the retardance
Rotation	$\begin{bmatrix} \cos\gamma & \sin\gamma \\ -\sin\gamma & \cos\gamma \end{bmatrix}$	axis rotation anticlockwise by γ from x

mits those components of an incident wave that are parallel to the x direction of its reference axis. The partial polariser is a practical implementation of the perfect polariser, and has independent transmission coefficients for components resolved parallel to the x and y directions of the polariser's reference axis. Ordinarily, $\tau_x >> \tau_y$ for a useful polariser, with the ratio τ_x/τ_y being a figure of merit, where large values are desirable. A pure retarder introduces a phase delay between the x and y components, but does not affect the magnitudes. Practical retarders ordinarily have some loss, and so a partial polariser matrix is included to account for this, although when the partial polariser matrix is used in this sense, $\tau_x \sim \tau_y$. Should the device be rotated with respect to the axes of the incident wave, it is necessary to include a rotation both before and after the device. An example is presented in due course.

In order to calculate the effect of series chain of devices, the matrices are multiplied together using standard matrix multiplication. For example, the effect of a partial polariser may be calculated thus (where the emergent electric fields are denoted by a superscript *):

$$\begin{bmatrix} E_x^* \\ E_y^* \end{bmatrix} = \begin{bmatrix} \tau_x & 0 \\ 0 & \tau_y \end{bmatrix} \begin{bmatrix} E_x \\ E_y \end{bmatrix}, \tag{A.27}$$

where the solution for x is

$$E_x^* = \tau_x E_{x_0} e^{j(\omega t - (2\pi z/\lambda) + \partial_x)}, \tag{A.28}$$

and for y:

$$E_y^* = \tau_y E_{y_0} e^{j(\omega t - (2\pi z/\lambda) + \partial_y)}, \tag{A.29}$$

.

If the optical element in question is birefringent, then the retardance may also be included:

$$\begin{bmatrix} E_x^* \\ E_y^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & e^{-j\Gamma} \end{bmatrix} \begin{bmatrix} \tau_x & 0 \\ 0 & \tau_y \end{bmatrix} \begin{bmatrix} E_x \\ E_y \end{bmatrix}, \tag{A.30}$$

yielding the solution in x as

$$E_x^* = \tau_x E_{x_0} e^{j(\omega t - (2\pi z/\lambda) + \partial_x)}, \tag{A.31}$$

as before, but for y it is

$$E_y^* = \tau_y E_{y_0} e^{j(\omega t - (2\pi z/\lambda) + \partial_y - \Gamma)}. \tag{A.32}$$

If the optical element is inclined anticlockwise from the x axis by an angle of γ, then the rotation matrices are also included

$$\begin{bmatrix} E_x^* \\ E_y^* \end{bmatrix} = \begin{bmatrix} \cos\gamma & -\sin\gamma \\ \sin\gamma & \cos\gamma \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & e^{-j\Gamma} \end{bmatrix} \begin{bmatrix} \tau_x & 0 \\ 0 & \tau_y \end{bmatrix} \begin{bmatrix} \cos\gamma & \sin\gamma \\ -\sin\gamma & \cos\gamma \end{bmatrix} \begin{bmatrix} E_x \\ E_y \end{bmatrix}. \tag{A.33}$$

This last equation may be more readily solved with the aid of numerical computing tools, since the algebraic expansion is unwieldy.

A.3 THE LINEAR HERMITIAN EIGENVALUE PROBLEM

In order to solve for the mode structure in a photonic crystal consisting of linear, lossless dielectric materials, in a source-free (no currents or charges) region, it is possible to begin with the macroscopic Maxwell's equations and arrive at a linear hermitian eigenvalue problem. The macroscopic Maxwell's equations may be derived from the microscopic equations according to the method of [206]. In cgs units, the equations are

$$\nabla \cdot \mathbf{B} = 0 \tag{A.34a}$$

$$\nabla \cdot \mathbf{D} = 4\pi\rho \tag{A.34b}$$

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \tag{A.34c}$$

$$\nabla \times \mathbf{H} = \frac{4\pi}{c}\mathbf{J} + \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t}, \tag{A.34d}$$

where $\mathbf{D}$ is the electric flux density, $\mathbf{B}$ is the magnetic flux density, $\mathbf{E}$ is the electric field, $\mathbf{H}$ is the magnetic field, ρ is the total charge, and $\mathbf{J}$ is the conduction (convection) current, with the use of the bold face denoting a vector quantity.

According to the derivation presented in Joannopoulos [107], in a source-free region with no charges or currents, it can be set that $\rho = \mathbf{J} = 0$. Assuming that the region is filled only with dielectric materials, and that any applied field strengths are small enough to assume linear behaviour, then higher order terms may be neglected in the power series

$$D_i = \sum_j \varepsilon_{ij} E_j + \sum_{jk} \chi_{ijk} E_j E_k + O(E^4), \tag{A.35}$$

that relates the components (E_i, D_i) of the electric and magnetic fields $(\mathbf{E}, \mathbf{H})$. Thus, at any position $\mathbf{r}$ within the dielectric, and for any angular frequency ω, the electric $\mathbf{E}(\mathbf{r}, \omega)$ and displacement $\mathbf{D}(\mathbf{r}, \omega)$ fields may be related by the scalar dielectric constant $\varepsilon(\mathbf{r}, \omega)$. Also, the frequency dependence of $\varepsilon(\mathbf{r}, \omega)$ may be neglected if a suitable value is assumed for the range of frequencies of interest. For lossy dielectrics, $\varepsilon(\mathbf{r})$ is complex. However, if it is assumed that only lossless dielectrics will be used, then $\varepsilon(\mathbf{r})$ may be considered to be purely real. Thus, $\mathbf{D}(\mathbf{r}) = \varepsilon(\mathbf{r})\mathbf{E}(\mathbf{r})$. At the high frequencies that are of interest in the study of photonic crystals (typically microwave through visible), the magnetic permeability of the dielectrics is close enough to unity for it to be set that $\mathbf{B} = \mathbf{H}$.

Now, Maxwell's equations become, as a function of position $\mathbf{r}$ and time t

$$\nabla \cdot \mathbf{H}(\mathbf{r}, t) = 0 \tag{A.36a}$$

$$\nabla \cdot \varepsilon(\mathbf{r})\mathbf{E}(\mathbf{r}, t) = 0 \tag{A.36b}$$

$$\nabla \times \mathbf{E}(\mathbf{r}, t) + \frac{1}{c}\frac{\partial \mathbf{H}(\mathbf{r}, t)}{\partial t} = 0 \tag{A.36c}$$

$$\nabla \times \mathbf{H}(\mathbf{r}, t) - \frac{\varepsilon(\mathbf{r})}{c}\frac{\partial \mathbf{E}(\mathbf{r}, t)}{\partial t} = 0. \tag{A.36d}$$

Since Maxwell's equations are linear, the field patterns can be treated separately from the time dependence by expanding the electric and magnetic fields into a set of sinusoidally harmonic modes. In phaser notation, a mode may be written as

$$\mathbf{H}(\mathbf{r}, t) = \mathbf{H}(\mathbf{r})e^{j\omega t}, \tag{A.37a}$$

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r})e^{j\omega t}. \tag{A.37b}$$

The time-independent field patterns $\mathbf{H}(\mathbf{r})$ and $\mathbf{E}(\mathbf{r})$ may be found by substitution into Equation A.36, under the conditions of given by the divergence equations:

$$\nabla \cdot \mathbf{H}(\mathbf{r}) = \nabla \cdot \mathbf{D}(\mathbf{r}) = 0, \tag{A.38}$$

where the time dependence is neglected because the divergence is assumed to be zero for all time. The implications of the zero-divergence condition are that (a) there are no sources or sinks of displacement or magnetic field, and (b) that the overall fields in the structure are constructed from waves that are transverse. Thus, for a plane wave $\mathbf{H}(\mathbf{r}) = \mathbf{a}\exp(i\mathbf{k} \cdot \mathbf{r})$, where $\mathbf{k}$ is the wavevector (not to be confused with the (scalar) wavenumber k or the grating vector $\mathbf{K}$), it is a requirement of Equation A.38 that $\mathbf{a} \cdot \mathbf{k} = 0$. Returning to the two curl equations, Equations A.36c & A.36d may be rewritten with the implied time dependence as

$$\nabla \times \mathbf{E}(\mathbf{r}) + \frac{i\omega}{c}\partial \mathbf{H}(\mathbf{r}) = 0, \tag{A.39a}$$

$$\nabla \times \mathbf{H}(\mathbf{r}) - \frac{i\omega}{c}\varepsilon(\mathbf{r})\mathbf{E}(\mathbf{r}) = 0. \tag{A.39b}$$

In order to decouple the equations so that they may be solved in terms of only $\mathbf{H}$ (to

allow the use of a Hermitian operator later on in the formation of the eigen problem), Equation A.39b maybe divided by $\varepsilon(\mathbf{r})$ and the electric field term rearranged to the right hand side of the equation

$$\frac{1}{\varepsilon(\mathbf{r})}\nabla \times \mathbf{H}(\mathbf{r}) = \frac{i\omega}{c}\mathbf{E}(\mathbf{r}). \tag{A.40}$$

and then the curl taken

$$\nabla \times \left(\frac{1}{\varepsilon(\mathbf{r})}\nabla \times \mathbf{H}(\mathbf{r})\right) = \frac{i\omega}{c}\nabla \times \mathbf{E}(\mathbf{r}) = 0, \tag{A.41}$$

and finally Equation A.39a is substituted to remove the dependence on the electric field, giving an equation in $\mathbf{H}(\mathbf{r}$

$$\nabla \times \left(\frac{1}{\varepsilon(\mathbf{r})}\nabla \times \mathbf{H}(\mathbf{r})\right) = \left(\frac{\omega}{c}\right)^2 \mathbf{H}(\mathbf{r}),. \tag{A.42}$$

In order to solve for the allowed modes that satisfy this relation, it is helpful to recast it into the format of an eigenvalue problem

$$\mathbf{A}v = \psi v \tag{A.43}$$

where $\mathbf{A}$ is a matrix, v is an eigenvector, and ψ is an eigenvalue. Thus, the left hand side of Equation A.42 may be identified as an operator Θ that operates on the magnetic field pattern $\mathbf{H}(\mathbf{r})$:

$$\Theta\mathbf{H}(\mathbf{r}) = \left(\frac{w}{c}\right)^2 \mathbf{H}(\mathbf{r}) \tag{A.44}$$

,where

$$\Theta\mathbf{H}(\mathbf{r}) \equiv \nabla \times \left(\frac{1}{\varepsilon(\mathbf{r})}\nabla \times \mathbf{H}(\mathbf{r})\right). \tag{A.45}$$

The field patterns of the harmonic modes are the eigenvectors $\mathbf{H}(\mathbf{r})$, and the frequency of these modes is proportional to the square root of the eigenvalues $(\frac{w}{c})^2$.

A.3.1 MIT Photonic Bands

The control file for the 1D simulation of a multilayer stack of alternating layers of GaAs ($\varepsilon = 13$) and air ($\varepsilon = 1$) layers, each layer of thickness $0.5a$, is as follows:

```
(set! num-bands 8)
(set! geometry-lattice (make lattice (size 1 no-size  no-size)))
(set! geometry (list (make block (center 0 0 0)
(material (make dielectric (epsilon 13)))
```

```
(size 0.5 0.5 0)
(e1 1 0 0 ) (e2 0 1 0 ) (e3 0 0 1 ))))
(set! k-points (list (vector3 -0.5 0 0) (vector3 0 0 0) (vector3 0.5 0 0)))
(set! k-points (interpolate 8 k-points))
(set! resolution 32)
(run)
```

The output frequencies are obtained with the following commands at the Linux command prompt:

```
linux> mpb multilayer.ctl > multilayer.out
linux> grep freqs multilayer.out > multilayer.dat
```

REFERENCES

[1] B. B. Hu and M. C. Nuss. Imaging with terahertz waves. *Opt. Lett.*, 20:1716–1720, 1995.

[2] D. M. Mittleman, R. H. Jacobsen, R. Neelamani, R. G. Baraniuk, and M. C. Nuss. Gas sensing using terahertz time-domain spectroscopy. *Appl. Phys. B*, 67:379 – 390, 1998.

[3] B. Ferguson and D. Abbott. De-noising techniques for terahertz responses of biological samples. *Microelectronics Journal*, 32:943 – 953, 2001.

[4] D. Woolard, R. Kaul, R. Suenram, A. Hight Walker, and A. Samuels. Terahertz electronics for chemical and biological warfare agent detection. *IEEE Microwave Theory and Techniques Society Digest*, 3:13 – 19, June 1999.

[5] D. Arnone, C. Ciesla, and M. Pepper. Terahertz imaging comes into view. *Physics World*, pages 35 – 40, April 2000.

[6] P. Siegel. Terahertz technology. *IEEE Trans. Microwave Theory Tech.*, 50(3):910 – 928, March 2002.

[7] R. Ulrich, T. J. Bridges, and M. A. Pollack. Variable metal mesh coupler for far infrared lasers. *Appl. Opt.*, 9(11):2511 – 2516, November 1970.

[8] R. D. Pollard. Guest editorial. *IEEE Trans. Microwave Theory Tech.*, 48(4):625, April 2000.

[9] D. H. Staelin, A. W. Morgenthaler, and J. A. Kong. *Electromagnetic Waves.* Prentice Hall Inc, Englewood Cliffs, New Jersey 07632, 1994.

[10] V. Illingworth, editor. *The Penguin Dictionary of Physics*, page 224. Market House Books Ltd, 3rd edition, 2000.

[11] S. Webb, editor. *The Physics of Medical Imaging.* Inst. of Physics Pub., 1988. ISBN: 0852743491.

[12] J. W. Waters. Submillimeter-wavelength heterodyne spectroscopy and remote sensing of the upper atmosphere. *Proc. IEEE*, 80(11):1679 – 1701, November 1992.

[13] G. Winnewisser, A. F. Krupnov, M. Yu. Tretyakov, M. Liedtke, F. Lewen, A. H. Saleck, R. Schieder, A. P. Shkaev, and S. V. Volokhov. Precision Broadband Spectroscopy in the Terahertz Region. *Journal of Molecular Spectroscopy*, 165:294 – 300, 1994.

[14] M. Walther, P. Plochocka, B. Fischer, H. Helm, and P Uhd Jepson. Collective vibrational modes in biological molecules investigated by terahertz time-domain spectroscopy. *Biopolymers*, 67:310 – 313, 2002.

[15] D. Grischkowsky, S. Keiding, M. van Exter, and Ch. Fattinger. Far-infrared time-domain spectroscopy with terahertz beams of dielectrics and semiconductors. *J. Opt. Soc. Am. B*, 7(10):2006 – 2015, October 1990.

[16] M.N. Afsar, I.I. Tkachov, and K.N. Kocharyan. Quasi-optical waveguide W-band spectrometer for precision dielectric measurement of absorbing materials. In *Conference Digest, Precision Electromagnetic Measurement*, pages 530–531, 1998.

[17] Committee on Free Electron Lasers and other Advanced Sources of Light, *et al.*. *Free Electron Lasers and other Advanced Sources of Light: Scientific Research Opportunities*. National Academy Press, Washington, DC, 1994.

[18] G. Mourou, C. V. Stancampiano, A. Antonetti, and A. Orszag. Picosecond microwave pulses generated with a subpicosecond laser driven semiconductor switch. *Appl. Phys. Lett.*, 39:295 – 296, August 1981.

[19] R. Heidemann, TH. Pfeiffer, and D. Jager. Optoelectronically pulsed slot-line antennas. *Electron. Lett.*, 19:316 – 317, April 1983.

[20] D.H. Auston, K.P. Cheung, and P.R. Smith. Picosecond photoconducting hertzian dipoles. *Appl. Phys. Lett.*, 45(3):284 – 286, August 1984.

[21] P. R. Smith, D. H. Auston, and M. C. Nuss. Subpicosecond photoconducting dipole antennas. *IEEE J. Quant. Electron.*, 24:255, 1988.

[22] M. B. Ketchen, D. Grischkowsky, T. C. Chen, C.-C. Chi, I. N. Duling III, N. J. Halas, J.-M. Halbout, J. A. Kash, and G. P. Li. Generation of subpicosecond electrical pulses on coplanar transmission lines. *Appl. Phys. Lett.*, 48(12):751 – 753, March 1986.

[23] D. R Grischkowsky, M. B. Ketchen, C.-C. Chi, I. N. Durling III, N. J. Halas, J.-M. Halbout, and P. G. May. Capacitance free generation and detection of

sub-picosecond electrical pulses on coplanar transmission lines. *IEEE J. Quant. Electron.*, 24(2):221 – 225, February 1988.

[24] A.P. DeFonzo, M. Jarwala, and C.R. Lutz. Transient response of planar integrated optoelectronic antennas. *Appl. Phys. Lett.*, 50:1155, 1987.

[25] A.P. DeFonzo and C.R. Lutz. Optoelectronic transmission and reception of ultrashort electrical pulses. *Appl. Phys. Lett.*, 51:212, 1987.

[26] C. Fattinger and D. R. Grischkowsky. Terahertz beams. *Appl. Phys. Lett.*, 54:490 – 492, 1989.

[27] Ch. Fattinger and D. Grischowsky. Point source terahertz optics. *Appl. Phys. Lett.*, 53:1480 – 1482, 1988.

[28] K. H. Yang, P. L. Richards, and Y. R. Shen. Generation of Far-Infrared Radiation by picosecond Light Pulses in $LiNbO_3$. *Appl. Phys. Lett.*, 19(9):320 – 322, 1971.

[29] M. Li, F. G. Sun, G. A. Wagoner, M. Alexander, and X.-C. Zhang. Measurement and analysis of terahertz radiation from bulk semiconductors. *Appl. Phys. Lett.*, 67:25, 1995.

[30] Q. Wu, T. D. Hewitt, and X.-C. Zhang. Two-dimensional electro-optic imaging of THz beams. *Appl. Phys. Lett.*, 69(8):1026 – 1028, August 1996.

[31] B. B. Hu, J. T. Darrow, X.-C. Zhang, D. H. Auston, and P. R. Smith. Optically steerable photoconducting antennas. *Appl. Phys. Lett.*, 56:886 – 888, 1990.

[32] X.-C. Zhang, B. B. Hu, J. T. Darrow, and D. H. Auston. Generation of femtosecond electromagnetic pulses from semiconductor surfaces. *Appl. Phys. Lett.*, 56:1011 – 1013, 1990.

[33] X.-C. Zhang and D. H. Auston. Optoelectronic measurement of semiconductor surfaces and interfaces with femtosecond optics. *J. Appl. Phys.*, 71:326–338, 1992.

[34] Q. Chen and X.-C. Zhang. Polarization modulation in optoelectronic generation and detection of terahertz beams. *Appl. Phys. Lett.*, 74(23):3435–3437, June 1999.

[35] Z. Jiang and X.-C. Zhang. Terahertz imaging via electro-optic effect. *IEEE Trans. Microwave Theory Tech.*, 47(12):2644 – 2650, 1999.

[36] Z. G. Lu, P. Campbell, and X.-C. Zhang. Free-space electro-optic sampling with a high-repetition-rate regenerative amplified laser. *Appl. Phys. Lett.*, 71(5):593 – 595, August 1997.

[37] U. Jepsen, C. Winnewisser, M. Schall, V. Schyja, S. R. Keiding, and H. Helm. Detection of THz pulses by phase retardation in lithium tantalate. *Phys. Rev. E*, 53:R3052 – R3054, 1996.

[38] A. Rice, Y. Jin, X. F. Ma, X.-C. Zhang, D. Bliss, J. Larkin, and M. Alexander. Terahertz optical rectification from $< 110 >$ zinc-blende crystals. *Appl. Phys. Lett.*, 64(11):1324 – 1326, March 1994.

[39] Bonvalet, M. Joffre, J. L. Martin, and A. Migus. Generation of ultrabroadband femtosecond pulses in the mid-infrared by optical rectification of 15 fs light pulses at 100 MHz repetition rate. *Appl. Phys. Lett.*, 67:2907 – 2909, 1995.

[40] D. W. van der Weide. Delta-doped Schottky diode nonlinear transmission lines for 480-fs, 3.5V transients. *Appl. Phys. Lett.*, 65(7):881 – 883, August 1994.

[41] J .H. Booske, C. L. Kory, D. Gallagher, V. Heinen, K. Kresicher, D. van der Weide, S. Limbach, P. Gustafson, W.-J. Lee, S. Gallagher, and K. Jain. Terahertz-regime, micro-VEDs: Evaluation of micromachine TWT conceptual designs. In *IEEE PPPS-2001 Conference Record (IEEE Cat. No. 01CH37255)*, Las Vegas, NV, USA, 17 – 22 June 2001.

[42] M. S. Shur and J.-Q. Lü. Terahertz sources and detectors using two-dimensional electronic fluid in high-electron-mobility transistors. *IEEE Trans. Microwave Theory Tech.*, 48(4):750 – 756, April 2000.

[43] V. M. Menon, W. D. Goodhue, A. S. Karakashian, A. Naweed, J. Plant, L. R. Ram-Mohan, A. Gatesman, V. Badami, and J. Waldman. Dual-frequency quantum cascade terahertz emitter. *Appl. Phys. Lett.*, 80(14):2454 – 2456, 2002.

[44] R. Köhler, A. Tredicucci, F. Beltram, H. E. Beere, E. H. Linfield, A. G. Davies, D. A. Ritchie, R. C. Iotti, and F. Rossi. Terahertz semiconductor-heterostructure laser. *Nature*, 417:156 – 159, May 9 2002.

[45] Y. Uzawa, A. Kawakami, S. Miki, and Z. Wang. Performance of all-NbN quasi-optical SIS mixers for the terahertz band. *IEEE transactions on Applied Superconductivity*, 11(1):183 – 186, March 2001.

[46] Y. Uzawa, S. Miki, Z. Wang, A. Kawakami, M. Kroug, P. Yagoubov, and E. Kollber. Performance of a quasi-optical NbN hot-electron bolometric mixer at terahertz frequencies. *Superconductor Science and Technology*, 15:141 – 145, 2002.

[47] M.C. Gaidis, H.M. Pickett, C.D. Smith, S.C. Martin, R.P. Smith, and P.H. Siegel. A 2.5THz receiver front end for spacebourne applications. *IEEE Trans. Microwave Theory Tech.*, 48(4):733 – 739, 2000.

[48] D. Leisawitz, W. Danchi, M. DiPirro, L. D. Feinberg, D. Gezari, M. Hagopian, W. D. Langer, J. C. Mather, H. Moseley Jr, M. Shao, R. F. Silverberg, J. Staguhn, M. R. Swain, H. W. Yorke, and Xiaolei Zhang. Scientific motivation and technology requirements for the SPIRIT and SPECS far-infrared/submillimeter space interferometers. *Proc. SPIE*, 4013:36 – 46, March 29 – 31 2000.

[49] N. C. Luhmann and W. A. Peebles. Instrumentation for magnetically-confined fusion plasma diagnostics. *Review of Scientific Instruments*, 55(3):279 – 331, March 1984.

[50] J. C. Wiltse. History of millimeter and submillimeter waves. *IEEE Trans. Microwave Theory Tech.*, MTT-32:1118 – 1127, September 1984.

[51] N. Gopalsami and A. C. Raptis. Remote detection of chemicals by millimeter wave spectroscopy. *Proceedings of the 4th SPIE international Conference on Millimeter and Submillimeter Waves Applications*, pages 254 – 265, July 1998.

[52] D. T. Petkie, T. M. Goyetter, R. P. A. Bettens, S. P. Belov, S. Albert, P. Helminger, and F. C. Delucia. A fast scan submillimeter spectroscopic technique. *Review of Scientific Instruments*, 68(4):1675 – 1683, April 1997.

[53] R. H. Jacobsen, D. H. Mittleman, and M. C. Nuss. Chemical recognition of gases and gas mixtures with terahertz waves. *Opt. Lett.*, 21(24):2011 – 2013, December 1996.

[54] D. Woolard *et al.*. The potential use of submillimeter-wave spectroscopy as a technique for biological warfare agent detection. *22nd Army Science Conference*, Dec 2000.

[55] D. Zimdars. Commerical T-ray systems accelerate imaging research. *Laser Focus World*, pages 91 – 96, July 2001.

[56] S. Thomsen, J. A. Pearce, and W.-F. Cheong. Changes in birefrigence as markers of thermal damage in tissues. *IEEE Transactions on Biomedical Engineering*, 36(12):1174–1179, December 1989.

[57] S. Jacques. Video imaging with polarized light finds skin cancer margins not visible to dermatologists. *Oregon Medical Laser Center News Etc*, February 1998.

[58] M. P. Rowe, E. N. Pugh Jr, J. S. Tyo, and N. Engheta. Polarisation-difference imaging: a biologically inspired technique for observation through scattering media. *Opt. Lett.*, 20(6):608 – 610, March 1995.

[59] S. Wang. Near-field polarization states and optical images in transmission mode through different surface structures. *J. Vac. Sci. Tech. B*, 17(6):2457–2461, 1999.

[60] T. Edwards. *Gigahertz and Terahertz Technologies for Broadband Communications*, chapter 1. Artech House, United Kingdom, 2000. ISBN: 1-58053-068-0.

[61] D. Turrettini and D. Young. Evaluating the Results of the UK 3G Spectrum License Auction. *http://emertech.wharton.upenn.edu/FordFoundation/Ford3GWireless.pdf*, 12 February 2002.

[62] D. C. Flanders. Submicrometer periodicity gratings as artificial anisotropic dielectrics. *Appl. Phys. Lett.*, 42(6):492 – 497, 1983.

[63] R. C. Enger and S. K. Case. Optical elements with ultrahigh spatial-frequency surface corrugations. *Appl. Opt.*, 22(20):3220–3228, 15 October 1983.

[64] L. H. Cescato, E. Gluch, and N. Streibl. Holographic quarterwave plates. *Appl. Opt.*, 29(22):3286–3290, 1 August 1990.

[65] D. R. S. Cumming and R. J. Blaikie. Variable polarisation compensator using artificial dielectrics. *Optics Communications*, 163:164–168, 1999.

[66] R. J. Blaikie, D. P. Taylor, and P. T. Gough. Multilevel differential polarization shift keying. *IEEE Transactions on Communications*, 45(1):95 – 102, 1997.

[67] R. J. Blaikie, M. M. Alkaisi, S. J. McNab, D. R. S. Cumming, R. Cheung, and D. G. Hasko. Nanolithography Using Optical Contact Exposure in the Evanescent Near Field. *Microelectronic Engineering*, 46:85 – 88, 1999.

[68] S. J. McNab, R. J. Blaikie, and M. M. Alkaisi. Analytic study of gratings patterned by evanescent near field optical lithography. *J. Vac. Sci. Tech.*, 18(6), November 2000.

[69] S. Ramo, J. R. Whinnery, and T. Van Duzer. *Fields and waves in communication electronics.* John Wiley and Sons, Inc., 3rd edition, 1994.

[70] J. M. Bell, G. H. Derrick, and R. C. McPhedran. Diffraction gratings in the quasi-static limit. *Optica Acta*, 29(11):1475 – 1489, 1982.

[71] D. Clarke and J. F. Grainger. *Polarised Light and Optical Measurement.* Pergamon Press, Oxford, 1971.

[72] W. L. Wolfe and G. J. Zeiss, editors. *The Infrared Handbook.* Office of Naval Research, Department of the Navy, Washington, DC, USA, 1978.

[73] E. Hecht and A. Zajac. *Optics.* Addison-Wesley Publishin Company, Reading, Massachusetts, 1974.

[74] M. Born and E. Wolf. *Principles of optics.* Cambridge University Press, Cambridge, 7th edition, 1999.

[75] Thorlabs, Inc., 435 Route 206, P.O. Box 366, Newton, NJ 07860-0366. *Fibre Polarization Controller*, 1998. http://www.thorlabs.com/Thorcat/0400/0482-D01.pdf.

[76] New Focus, Inc. Polarization and polarization control. *Application Note 3*, 1993.

[77] D. Clarke and J. F. Grainger. *Polarised Light and Optical Measurement*, page 105. Pergamon Press, Oxford, 1971.

[78] D. Clarke and J. F. Grainger. *Polarised Light and Optical Measurement*, page 103. Pergamon Press, Oxford, 1971.

[79] E. D. Palik, editor. *The handbook of optical constants of solids*, volume 3. Academic Press Limited, London, UK, 1998.

[80] E. Hecht and A. Zajac. *Optics*, pages 312–313. Addison-Wesley Publishin Company, Reading, Massachusetts, 1974.

[81] G. Lammel, S. Schweizer, and P. Renaud. MEMS infrared gas spectrometer based on a porous silicon tunable filter. *IEEE Conference* $0-7803-5998-4/01$, page 578, 2001.

[82] P.B. Clapham and M.C. Hutley. Reduction of lens reflexion by the 'moth eye' principle. *Nature*, 244:281, 1973.

[83] D.H. Raguin and G.M. Morris. Analysis of antireflection-structured surfaces with continuous one-dimensional surface profiles. *Appl. Opt.*, 32(14):2582 – 2598, 1993.

[84] B. Macleod and G. Sonek. Motheye surfaces reflect little light. *Laser Focus World*, 35(8):109 – 114, 1999.

[85] C. G. Bernhard. Structural and functional adaption in a visual system. *Endeavour*, 26:79 – 84, 1967.

[86] S.M. Rytov. Electromagnetic properties of a finely stratified medium. *Soviet Physics JETP*, 2(3):466–475, May 1956.

[87] S. Ramo, J. R. Whinnery, and T. Van Duzer. *Fields and waves in communication electronics*, page 293. John Wiley and Sons, Inc., 3rd edition, 1994.

[88] W. Thornburg. The form birefringence of lamellar systems containing three or more components. *Journal of Biophysical and Biochemical Cytology*, 3(3):413 – 419, 1997.

[89] M. Rosser, editor. *Formulae and Tables.* School of Engineering, University of Auckland, Auckland, New Zealand, 2nd edition, January 1990.

[90] P. Lalanne and D. Lemercier-Lalanne. Depth dependance of the effective properties of subwavelength gratings. *J. Opt. Soc. Am. A*, 14(2):450–458, February 1997.

[91] Y. Ono, Y. Kimura, Y. Ohta, and N. Nishida. Antireflection effect in ultrahigh spatial-frequency holographic relief gratings. *Appl. Opt.*, 26(6):1142 – 1146, 15 March 1987.

[92] V. Kettunen and F. Wyrowski. Reflection-mode phase retardation by dielectric gratings. *Optics Communications*, 158(1 – 6):41 – 44, 1998.

[93] L.-M. Li. Two-dimensional photonic crystals: candidate for wave plates. *Appl. Phys. Lett.*, 78(22), May 2001.

[94] E. Yablonovitch. Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.*, 58(20):2059 – 2062, May 1987.

[95] S. John. Strong localisation of photons in certain disordered dielectric superlattices. *Phys. Rev. Lett*, 58(23):2486 – 489, June 1987.

[96] E. Yablonovitch, T. J. Gmitter, R. D. Meade, K. D. Brommer, A. M. Rappe, and J. D.Joannopoulos. Donor and acceptor modes in photonic band structure. *Phys. Rev. Lett.*, 67:3380, 1991.

[97] K. M. Ho, C. T. Chan, and C. K. Soukalis. Existence of a photonic bandgap in periodic dielectric structures. *Phys. Rev. Lett.*, 65(25):3152 – 3155, December 1990.

[98] S.-Y. Lin, E. Chow, V. Hietala, P. R. Villeneuve, and J. D. Joannopoulos. Experimental Demonstration of Guiding and Bending of Electromagnetic Waves in a Photonic Crystal. *Science*, 282:pp. 274 – 276, 9 October 1998.

[99] K. Hosomi and T. Katsuyama. A dispersion compensator using coupled defects in a photonic crystal. *IEEE J. Quant. Electron.*, 38(7):825 –829, July 2002.

[100] S. John and R. Rangarajan. Optimal structures for classical wave localization: an alternative to the Ioffe-Regel criterion. *Phys. Rev. B*, 38(14):10101, November 1988.

[101] K. M Leung and Y. F. Liu. Full vector calculation of photonic band structures in face centered cubic dielectric media. *Phys. Rev. Lett.*, 65(12):2646 – 2649, November 1990.

[102] S. Satpathy *et al.*. Theory of photon bands in three-dimensional periodic dielectric structures. *Phys. Rev. Lett.*, 64(11):1239 – 1242, March 1990.

[103] E. Yabonlonovitch and T. J. Gmitter. Photonic band structures: the face-centered-cubic case. *Phys. Rev. Lett.*, 63(18):1950 – 1953, October 1989.

[104] Z. Zhang *et al.*. Electromagnetic wave propagation in periodic structures: Bloch wave solution of Maxwell's equation. *Phys. Rev. Lett.*, 65(21):2650 – 2653, November 1990.

[105] J. Maddox. Photonic band-gaps bite the dust. *Nature*, 348:481, December 1990.

[106] H. S. Sözüer *et al.*. Photonic bands: convergence problems with the plane wave method. *Phys. Rev. B.*, 45(24):13962 – 13973, June 1992.

[107] John D. Joannopoulos, Robert D. Meade, and Joshua N. Winn. *Photonic Crystals molding the flow of light.* Princeton University Press, Princeton, New Jersey, USA, 1995.

[108] Steven G. Johnson and J. D. Joannopoulos. Block-iterative frequency-domain methods for Maxwell's equations in a planewave basis. *Opt. Express*, 8(3):173–190, 2001.

[109] J. B. Pendry. Photonic band structures. *J. Mod. Opt.*, 41(2):209 – 229, 1994.

[110] P. M. Bell, J. B. Pendry, L. Martin Moreno, and A. J. Ward. A program for calculating photonic band structures and transmission coefficients of complex structures. *Computer Physics Communications*, 85(2):306–322, February 1995.

[111] A. L. Reynolds and J. M. Arnold. Interleaving two-dimensional lattices to create three-dimensional photonic bandgap structures. *IEE Proc. Optoelectron.*, 145(6):436 – 440, December 1998.

[112] L. Zhang *et al.*. An efficient finite-element method for the analysis of photonic band gap materials. *IEEE Microwave Theory Tech.-Soc. Digest*, pages 1703 – 1706, 1999.

[113] H. Y. Yang. Characteristics of guided and leaky waves on multilayer thin-film structures with planar material gratings. *IEEE Trans. Microwave Theory Tech.*, pages 428 – 435, March 1997.

[114] K. M. Ho, C. T. Chan, C. K. Soukalis, R. Biswas, and M. Sigalas. Photonic band gaps in three dimensions: new layer-by-layer periodic structures. *Solid State Communications*, 89(5):413 – 416, 1994.

[115] H. S. Sözüer *et al.*. Photonic bands: simple-cubic lattice. *J. Opt. Soc. Am. B*, 10(2):296 –302, February 1993.

[116] E. Özbay, E. Michel, G. Tuttle, R. Biswas, M. Sigalas, and K.-M. Ho. Micromachined millimeter-wave photonic band-gap crystals. *Appl. Phys. Lett*, 64(16):2059 – 2061, April 1994.

[117] E. Özbay, E. Michel, G. Tuttle, R. Biswas, K.-M. Ho, J. Bostak, and D. M. Bloom. Terahertz spectroscopy of three-dimensional photonic band-gap crystals. *Opt. Lett.*, 19(15):1155 – 1157, August 1994.

[118] A. Chelnokov, S. Rowson, J.-M. Lourtioz, L. Duvillaret, and J.-L. Courtioz. Terahertz characterisation of mechanically machined 3d photonic crystal. *Electron. Lett.*, 33(23):1981 – 1983, November 1997.

[119] S. Y. Lin, J. G. Fleming, D. L. Hetherington, B. K. Smith, R. Biswas, K. M. Ho, M. M. Sigalas, W. Zubryzycki, S. T. Kurtz, and J. Bur. A three-dimensional photonic crystal operating at infrared wavelengths. *Nature*, 394:251 – 253, July 1998.

[120] E. Yablonovitch, T. J. Gmitter, R. D. Meade, K. D. Brommer, A. M. Rappe, and J. D.Joannopoulos. Donor and acceptor modes in photonic band structure. *Phys. Rev. Lett*, 67(24):3380 – 3383, December 1991.

[121] D. R. Smith, R. Dalichaouch, N. Kroll, S. Schultz, S. L. McCall, and P. M. Platzman. Photonic band structure and defects in one and two dimensions. *J. Opt. Soc. Am. B*, 10(2):314 – 321, February 1993.

[122] P. R. Villeneuve, S. Fan and J. D. Joannopoulos. Microcavities in photonic crystals: mode symmetry, tunability and coupling efficiency. *Phys. Rev. B*, 54(11):7837 – 7842, September 1996.

[123] S. Fan, P.R. Villeneuve, and J. D. Joannopoulos. Large omni-directional band gaps in metallodielectric photonic crystals. *Phys. Rev. B*, 54(16):11 245 – 11251, October 1996.

[124] F. Gadot, A. de Lustrac, J.-M. Lourtioz, T. Brillat, A. Ammouche, and E. Akmansoy. High-transmission defect modes in two-dimensional metallic photonic crystals. *J. Appl. Phys.*, 85(12):8499 – 8501, June 1999.

[125] F. Gadot, T. Brillat, E. Akmonsoy, and A. de Lustrac. New type of metallic photonic bandgap material suitable for microwave applications. *Electron. Lett.*, 36(7):640 – 641, March 2000.

[126] W. Y. leung, G. Tuttle, M. M. Sigalas, R. Biswas, K. M. Ho, and C. M. Soukoulis. Optimizing the Q value in three-dimensional metallic photonic band gap crystals. *J. Appl. Phys.*, 84(8):4091 – 4095, October 1998.

[127] D. W. Porterfield, J. L. Hesler, R. Densing, E. R. Mueller, T. W. Crowe, and R. M. Weikle II. Resonant metal-mesh bandpass filters for the far-infrared. *Appl. Opt.*, 33(25):6046 – 6052, September 1994.

[128] B. Temulkuran, M. Bayindir, E. Ozbay, J.P. Kavanaugh, M. M. Sigalas and G. Tuttle. Quasimetallic silicon micromachined photonic crystals. *Appl. Phys. Lett.*, 78(3):264 – 266, January 2001.

[129] S. Kim and V. Gopalan. Strain-tunable photonic bandgap crystals. *Appl. Phys. Lett.*, 78(20):3015, 2001.

[130] E. Yablonovitch. Liquid vs photonic crystals. *Nature*, 401:539 – 541, October 1999.

[131] H. Koops, W. Dultz, and M. Eich. Method for the mechanical stabilization and for tuning a filter having a photonic crystal structure. *United States Patent*, October 26 1999. Patent Number 5,973,823.

[132] J.-M. Lourtioz, A. de Lustrac, F. Gadot, A. Chelnokov, T. Brillat, A. Ammouche, J. Danglot, O. Vanbésien, and D. Lippens. Toward controllable photonic crystals for centimeter and millimeter-wave devices. *J. Lightwave Tech.*, 17(11):2025 – 2031, November 1999.

[133] P. Bondavalli, R. LeDantec, and T. Benyattou. Influence of the involuntary underetching on the mechanical properties of tunable Fabry-Pérot filters for optical communications. *J. Microelectromech. Syst.*, 10(2):198, 2001.

[134] H. Alause, J.P. Malzoc, F. Grasdepot, V. Nouaze, J. Hermann, and W. Knap. Micromachined optical tunable filter for long term gas stability sensors. *IEE Proc. Optoelectronics*, 144(5), October 1997.

[135] J. P. Gianvittorio, J. Zendejas, Y. Rahmat-Samii, and Jack W. Judy. Mems enabled reconfigurable frequency selective surfaces: Design, simulation, fabrication, and measurement. *APS Int'l Symposium*, 2:pp. 404–407, 2002.

[136] Wayne M. Moreau. *Semiconductor Lithography Principles, Practices and Materials.* Micro Devices physics and fabrication technologies. Plenum Press, New York, 1988.

[137] Shipley Far East Ltd. 3F Sanbancho UF Building, 6-3 Sanbancho, Chiyoda-Ku, Tokyo 102-0075, Japan. http://www.shipley.com, Accessed 30 May 2002.

[138] (USA) Clariant Corporation. Business Unit Electronic Materials, 70 Meister Avenue, Somerville, NJ 08876, USA. http://www.clariant.com, Accessed 30 May 2002.

[139] H. Chen. Nanolithography using reactive ion etched silicon nitride phase shifting masks. Master's thesis, University of Canterbury, February 2000.

[140] K. K. Ma. Micro-machining of monolithic silicon waveguides for terahertz frequencies based on rectangular waveguides. Master's thesis, University of Canterbury, Christchurch, New Zealand, August 2000.

[141] K.R. Williams and R.S. Muller. Etch rates of micromachining processing. *IEEE J. of Microelectromechanical Systems*, 5(4):256–269, December 1996.

[142] S.M. Sze. *Semiconductor Devices: Physics and Technology.* Wiley, New York, 1985.

[143] K. E. Petersen. Silicon as a mechanical material. *Proc. IEEE*, 70:420–457, May 1982.

[144] K.E. Bean. Anistropic etching of silicon. *IEEE Trans. Electron Devices*, ED-25(10):1185–1193, October 1978.

[145] G. T. A. Kovacs and K. E. Petersen N. I. Maluf. Bulk micromachining of silicon. *Proc. IEEE*, 86(8):1536–1551, August 1998.

[146] H. Seidel, L.Csepregi, A.Heuberger, and H.Baumgärtel. Anisotropic etching of crystalline silicon in alkaline solutions I: Orientation dependence and behaviour of passivation layers. *J. Electrochem Soc.*, 137(11):3612–3626, November 1990.

[147] K. K. Ma, T. D.Drysdale, R. J. Blaikie, and D. R. S.Cumming. Novel silicon bulk micromachining process for submillimetre rectangular waveguide fabrication. In *Proc. SPIE Microelectronic and MEMS Technologies*, pages 372 – 379, Edinburgh, Scotland, May 30 - June 1 2001.

[148] E. Bassous. Fabrication of novel three-dimensional microstructures by anisotropic etching of (100) and (110) silicon. *IEEE Trans. On Electron Devices*, ED-25(10):1179–1185, October 1978.

[149] D.L.Kendall. A new theory for the anistropic etching of silicon and some underdeveloped chemical micromachining concepts. *J. Vacuum Sicenece Technology A*, 8(4):3598–3605, July 1990.

[150] A. J. Nijdam, E. van Veenendaal, H. M. Cuppen, J. van Suchtelen, M. L. Reed, J. G. E. Gardeniers, W. J. P. van Enckevort, E. Vlieg, and M. Elwenspoek. Formation and stabilization of pyramidal etch hillocks on silicon in anisotropic etchants: Experiments and Monte Carlo simulation. *J. Appl. Phys.*, 89(7):4113–4123, April 2001.

[151] H.Han, R.Boudreau, Terry Bowen, S.Tan, and M.L. Reed. Etching defects on koh etched silicon - implementation of silicon bench technology for low cost packaging. *Proc. 44th Electronic Components and Technology Conf.*, pages 625–631, 1994.

[152] University of Illinois at Urbana-Champaign. Anisotropic Crystalline Etch Simulation (ACES) computer program for Windows 95/NT. http://mass.micro.uiuc.edu/research/completed/aces/pages/home.html, Accessed 1999.

[153] Z. Zhu and C. Liu. Simulation of anisotropic crystalline etching using a continuous cellular automata algorithm. *Journal Computer Modeling for Engineering and Science*, 1(1):11–19, 2000.

[154] D.M. Allen and I.A. Routledge. Anisotropic etching of silicon: a model diffusion controlled reaction. *Proc. IEEE*, 130(2):49–56, 1983.

[155] D.R. Ciarlo. A latching accelerometer fabricated by the anisotropic etching of (110) oriented silicon wafers. *Journal of Micromechanics and Microengineering*, 2:10–13, 1992.

[156] Mattias Vangbo and Ylva Bäcklund. Precise mask alignment to the crystallographic orientation of silicon wafers using wet etching. *Journal of Micromechanics and Microengineering*, 6:279–284, 1996.

[157] G. Ensell. Alignment of mask patterns to crystal orientation. In *The 8th International Conference on Solid-State Sensors and Actuators, and Eurosensors*, pages 186–189, June 1995.

[158] G. Turner of Industrial Reseach Limited Gracefield Research Facility. Pers. Comm., Oct 1999.

[159] E. Da Silva, editor. *High frequency and microwave engineering.* Butterworth-Heinemann, Oxford, 2001.

[160] Cascade Microtech Ltd. Wincal version 2.2 software for vector network analysers. http://www.cascademicrotech.com, Accessed February 2002.

[161] W. L. Barrow and L. J. Chu. Theory of the electromagnetic horn. *Proceedings of the Institute of Radio Engineers*, 27:51–64, Jan 1939.

[162] C. A. Balanis. *Antenna Theory analysis and design.* John Wiley & Sons, Inc., Singapore, 1982.

[163] D. R. S. Cumming. Photographs of the W-band test platform at the University of Glasgow, 1999.

[164] Flann Microwave Instruments. *Metrology Bronze Grade Waveguide Calibration Kit Model 27703 Serial No. 22.*

[165] John H. Davies. *The physics of low-dimensional semiconductors.* Cambridge University Press, 1998.

[166] Allen Taflove and Susan C. Hagness. *Computational Electrodynamics: the finite difference time domain method.* Artech House, Boston, second edition, 2000.

[167] Roger F. Harrington. *Field Computation by Moment Methods.* Macmillan, New York, 1968.

[168] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. T. Flannery. *Numerical Recipies in C - The art of scientific computing.* Cambridge University Press, Cambridge, second edition, 2002.

[169] E. K. Miller, L. Medgyesi-Mitschang, and Edward H. Newman, editors. *Computational Electromagnetics - Frequency Domain Method of Moments.* IEEE Press, New York, 1992.

[170] D. Fluckiger. Antennasolver. NEC2-based antenna solver software, P.O. Box 353 Allen TEXAS 75013, http://www.gsolver.com. http://www.gsolver.com.

[171] National Technical Information Service (U.S. Department of Commerce), 5285 Port Royal Road, Springfield, VA 22161. *Numerical Electromagnetics Code - Method of Moments.* http://www.ntis.gov.

[172] Hewlett Packard. High frequency structure simulator. Reference Manual, 2000. p. A10.

[173] Hewlett Packard. High frequency structure simulator. Reference Manual, 2000. p. A11.

[174] C. Hafner. *The Generalized Multipole Technique for Computational Electromagnetics.* Artech House, Boston, 1990.

[175] C. Hafner. Multiple multipole program computation of periodic structures. *J. Opt. Soc. Am. A*, 12(5):1057–1067, May 1995.

[176] M. G. Moharam, D. A. Pommet, and Eric B. Grann. Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach. *J. Opt. Soc. Am. A*, 12(5):1077 – 1086, 1995.

[177] The Gsolver FAQ. http://www.gsolver.com/faq.html, November 2002.

[178] Grating Solver Development Company. Gsolver4.12. http://www.gsolver.com.

[179] L. I. Goray. Modified integral method for weak convergence problems of light scattering on relief grating. *SPIE Proceedings*, 4291:1–12, 2001.

[180] P. Lalanne and J.-P. Hugonin. High-order effective-medium theory of subwavelength gratings in classical mounting: application to volume holograms. *J. Opt. Soc. Am. A*, 15(7):1843–1851, July 1998.

[181] M. van Exter and D. Grischkowsky. Carrier dynamics of electrons and holes in moderately doped silicon. *Phys. Rev. B*, 41(17):12140 – 12149, June 1990.

[182] M. N. Afsar, I. I. Tkachov, and K. N. Kocharyan. A novel W-band spectrometer for dielectric measurements. *IEEE Trans. Microwave Theory and Tech.*, 48(12):2637 – 2643, December 2000.

[183] S. Arscott, F. Garet, P. Mounaix, L. Duvillaret, J.-L. Coutaz, and D. Lippens. Terahertz time-domain spectroscopy of films fabricated from SU-8. *Electron. Lett.*, 35(3):243 – 244, 1999.

[184] Bernd Loechel. Thick-layer resists for surface micromachining. *J. Micromech. Microeng.*, 10:108 – 115, 2000.

[185] M.C. Wanke, O. Lehmann, K. Müller, Q. Wen, and M. Stuke. Laser rapid prototyping of photonic band-gap microstructures. *Science*, 275:1284 – 1286, February 1997.

[186] B. Temelkuran, H. Altug, and E. Ozbay. Experimental investigation of layer-by-layer metallic photonic crystals. *IEE Proc. Optoelectronics*, 145(6):409 – 414, 1998.

[187] E. Ozbay, B. Temelkuran, M. Sigalas, G. Tuttle, C. M. Soukalis, and K.M. Ho. Defect structures in metallic photonic crystals. *Appl. Phys. Lett.*, 69(25):3797 – 3799, 1996.

[188] B. Temelkuran, E. Ozbay, M. Sigalas, G. Tuttle, C.M. Soukoulis, and K.M. Ho. Reflection properties of metallic photonic crystals. *Appl. Phys. A*, 66, 363 – 365 1998.

[189] B. Temelkuran, Mehmet Bayinder, E. Ozbay, J. P. Kavanaugh, M. M. Sigalas, and G. Tuttle. Quasimetallic silicon micromachined photonic crystals. *Appl. Phys. Lett.*, 78(3):264 – 266, 2001.

[190] A. R. Reynolds. Translight photonic crystal analysis code. http://www.elec.gla.ac.uk/groups/opto/photoniccrystal/Software/Software%Main.htm, November 2002.

[191] T. V. Pistor, K. Adam, and A. Neureuther. Rigorous simulation of mask corner effects in extreme ultraviolet lithography. *J. Vac. Sci. Tech. B*, 16(6):3449–3455, Nov 1998.

[192] Remcom XFDTD. http://www.remcom.com, November 2002.

[193] ISE EMLAB. http://www.ise.ch/products/index.html, November 2002.

[194] Allen Taflove and Susan C. Hagness. *Computational Electrodynamics: the finite difference time domain method*, page 314. Artech House, Boston, second edition, 2000.

[195] Allen Taflove and Susan C. Hagness. *Computational Electrodynamics: the finite difference time domain method*, pages 133 – 140. Artech House, Boston, second edition, 2000.

[196] John G. Proakis and Dimitris G. Manolakis. *Digital Signal Processing: Principles, Algorithms and Applications.* Prentice Hall, 3rd edition, 1995.

[197] A.L. Reynolds, H. Chong, I.G. Thayne, P.J.I. de Maagt, and J.M. Arnold. Transmission response for in-plane and out of plane propagation for a 2D photonic crystal with a planar cavity. *Synthetic Metals*, 116:433 – 437, 2001.

[198] W. Nakagawa, P.-C. Sun, C.-H. Chen, and Y. Fainman. Wide-field-of-view narrow-band spectral filters based on photonic crystal nanocavities. *Opt. Lett.*, 27:191 – 193, 2002.

[199] D.F. Sievenpiper, M.E. Sickmiller, and E. Yablonovitch. 3D wire mesh photonic crystals. *Phys. Rev. Lett.*, 76(14):2480 – 2483, 1996.

[200] W. Nakagawa and Y. Fainman. Tunable optical nanocavity based on modulation of near-field coupling between subwavelength periodic nanostructures. In *Diffractive Optics & Micro Optics*, Tucson, Arizona, 3 – 6 June 2002.

[201] E. F. Crawley and M. A. de Luis. Embedded piezoelectric structure and control. *U.S. Patent No. US4849668*, May 1987.

[202] R. J. Blaikie, T. Drysdale, H. M. H. Chong, I. G. Thayne, and D. R. S. Cumming. Experimental realization of wide-field-of-view spectral filters baed on large-period defects in photonic crystals. *Opt. Lett.*, To be submitted.

[203] Picometrix, Inc. http://www.picometrix.com, 2925 Boardwalk, Ann Arbor, MI 48104-6765, U.S.A.

[204] TeraView Ltd. http://www.teraview.co.uk, Cambridge, U.K.

[205] J. D. Kraus. *Electromagnetics.* McGraw-Hill International Book Company, London, 3rd edition, 1984.

[206] J. D. Jackson. *Classical Electrodynamics.* John Wiley & Sons, New York, 1962.